Spatial Determinants of Entrepreneurship in India

Spatial Determinants of Entrepreneurship in India

Ejaz Ghani, William R. Kerr and Stephen O'Connell

August 2013

Regional Studies, forthcoming

Abstract: We analyze the spatial determinants of entrepreneurship in India in the manufacturing and services sectors. Among general district traits, quality of physical infrastructure and workforce education are the strongest predictors of entry, with labor laws and household banking access also playing important roles. We also find extensive evidence of agglomeration economies among manufacturing industries. In particular, supportive incumbent industrial structures for input and output markets are strongly linked to higher establishment entry rates. In comparison to the U.S., regional conditions in India play a stronger relative role for the spatial patterns of entrepreneurship compared to incumbent industry locations.

Keywords: Entrepreneurship, agglomeration, development, India, South Asia.

JEL Classification: L10, L26, L60, L80, M13, O10, R00, R10, R12

Author institutions and contact details: Ghani: World Bank, Eghani@worldbank.org; Kerr: Harvard University, Bank of Finland, and NBER, wkerr@hbs.edu; O'Connell: World Bank and CUNY Graduate Center, soconnell@gc.cuny.edu.

Acknowledgments: We thank Ahmad Ahsan, Mehtabul Azam, Rajeev Dehejia, Gilles Duranton, Arti Grover, Michael Fritsch, Lakshmi Iyer, Henry Jewell, Arvind Panagariya, David Storey, Hyoung Gun Wang, and three anonymous referees for helpful comments on this work. We thank the World Bank's South Asia Labor Flagship team for providing the primary datasets used in this paper. We are particularly indebted to Shanthi Nataraj for sharing her wisdom regarding the industrial survey data. Funding for this project was provided by World Bank and Multi-Donor Trade Trust Fund. The views expressed here are those of the authors and not of any institution they may be associated with.

I. Introduction

Many policy makers want to encourage entrepreneurship given its perceived role in economic growth and development.[1] The importance of this factor has led to extensive recent research on regional traits associated with entrepreneurship. Multiple studies consider advanced economies, but there is very little empirical evidence for developing countries. This lack of research hampers the effectiveness of policy: for example, the roles that education or infrastructure play in entry in the U.S. may be quite different from a setting where illiteracy and lack of roads and sanitation continue to hamper development.

Audretsch et al. (2012) emphasize the local nature of entrepreneurship determinants. We investigate these questions for manufacturing and services in India. Within these two industry groups, we also compare the organized and unorganized sectors. We quantify the traits of districts that systematically predict stronger entry levels. Several important themes emerge from our study. First, education levels and local infrastructure access are the most prominent local traits linked to entrepreneurship across all sectors. Second, local industrial conditions—the links that form across industries within a district—play an even stronger role in predicting entry within specific district-industries than the general district-level traits. Finally, in comparison to the U.S., we find that India's economic geography is still taking shape. At such an early point and with industrial structures not entrenched, there is room for policy to have substantial impact by shaping where industries plant their roots.

Our study makes several contributions to the literature. We are among the first studies to quantify the spatial determinants of entrepreneurship in India. Moreover, we move beyond manufacturing to consider services, and we compare the organized and unorganized sectors. The latter analyses of the unorganized sector are among our most important contributions given the limited study of the informal economy previously and its substantial importance for India and other developing economies. More broadly, we are among the first studies to apply the incumbent industrial structures frameworks of Glaeser and Kerr (2009) to a developing economy, providing insights into how agglomeration economies resemble and differ from each other. More research on agglomeration economies and entrepreneurship in developing countries is important for urban and development economics going forward.[2]

[1] High rates of local entrepreneurship are linked to stronger subsequent job growth for regions in several countries (e.g., Fritsch 2008, Ghani et al. 2011, Glaeser et al. 2012). Mueller et al. (2008) caution, however, about sweeping statements given the substantial heterogeneity in the British experience, where job growth of regions depended strongly on the types of entrepreneurs entering and the initial conditions of the regions. Baumol (1990) also highlights how the positive or negative role of entrepreneurship depends upon the incentives in society.

[2] In contemporaneous work, Mukim (2011) examines spatial entry patterns for India's unorganized sector. Our working paper discusses similarities and differences between our studies. Other related work includes Drucker

Identifying local conditions that encourage entrepreneurship and acting upon them is essential to foster economic growth. Figure 1 shows that entrepreneurship rates are lower in South Asia than what its stage of development would suggest. Effective entrepreneurship will play a key role in job growth for India, the development of a strong manufacturing base (Fernandes and Pakes 2010), and the transition of people out of subsistence living and the informal sector. Khanna (2008) emphasizes entrepreneurship for India's future, and reallocation can help close India's productivity gap (e.g., Hsieh and Klenow 2009).[3]

II. Spatial Entrepreneurship Rates in India

We measure entrepreneurship as the presence of young establishments. Our primary measure, which we can consistently observe across all of our datasets, is whether an establishment is less than three years old. For the organized manufacturing sector, we can also measure establishments in their first year of existence, and we find very similar results with this approach. Incumbent establishments, which are used to model existing activity in the district-industry, are firms that are three years old or more. We principally define entry measures through employment in young establishments, and we look at counts of entering establishments in robustness checks.[4]

We employ establishment-level surveys of manufacturing and service enterprises carried out by the Government of India. Our manufacturing data are taken from surveys conducted in fiscal years 2005-06; services sector data come from 2001-02. While these surveys were conducted over two fiscal years, we refer below to the initial year only. An unpublished appendix is available from the authors upon request. This appendix lists data sources and years employed, and we describe below additional information included in this appendix. Nataraj (2009), Kathuria et al. (2010), Hasan and Jandoc (2010), and Dehejia and Panagariya (2010) provide detailed overviews of similar databases.

---

and Feser (2007, 2012), Acs and Varga (2005), Ardagna and Lusardi (2008), Rosenthal and Strange (2010), Delgado et al. (2010), and Calá et al. (2013).

[3] Parker (2009a) provides a complete review of the entrepreneurship literature, and Storey (1994) and Storey and Greene (2010) give an overview of small businesses and their connections to entrepreneurship specifically. Deichmann et al. (2008) survey prior work on firm locations in developing economies.

[4] Our data combine single-unit start-ups with expansion facilities of multi-unit firms. We can, to some degree, separate entry of multi-unit firms within organized manufacturing, although this distinction is not comprehensively available for all plants. With the splits available, we find very similar results when modeling single-unit entry rates. These splits are not possible for the unorganized sectors and services. A major development limitation for India is the growth and replication of successful initial businesses (e.g., Hsieh and Klenow 2009). From this perspective, many policy makers are equally concerned about encouraging entry of expansion establishments. Our working paper provides an extended discussion about our measures of entrepreneurship and alternative approaches. We also return to this discussion when considering the entrant size distribution.

The distinction between organized and unorganized sectors relate to establishment size. In manufacturing, the organized sector is comprised of establishments with more than ten workers if the establishment uses electricity. If the establishment does not use electricity, the threshold is 20 workers or more. These establishments are required to register under the Factories Act of 1948. The unorganized manufacturing sector is, by default, comprised of establishments which fall outside the scope of the Factories Act.

Service establishments, regardless of size or other characteristics, are not required to register and thus are all officially unorganized. There are various approaches to comparably differentiate small-scale, autonomous establishments from the larger employers that constitute the organized sector, as generally defined. We assign services establishments with fewer than five workers and/or listed as an "own-account enterprise" (OAE) to the unorganized sector. OAE enterprises are firms that do not employ any hired worker on a regular basis. The choice of five employees as the size cutoff recognizes that average establishment size in services is significantly smaller than in manufacturing. Using this demarcation, the organized sector makes up approximately 25% of employment in both manufacturing and services.

The organized manufacturing sector is surveyed by the Central Statistical Organisation (CSO) every year through the Annual Survey of Industries (ASI), while unorganized manufacturing and services establishments are separately surveyed by the National Sample Survey Organisation (NSSO) at approximately five-year intervals. Establishments are surveyed with state and four-digit National Industry Classification (NIC) stratification. For organized manufacturing, the business register described above forms the basis for the sampling frame. Establishments are notified if they fall into the sampled frame and are required by law to complete and return the survey questionnaire, and the CSO investigates cases of non-response (typically closed plants). For the services and unorganized manufacturing sector, India's Economic Census comprises the basis for the sampling frame and stratification procedures. Establishments falling into the sample are then surveyed by government enumerators.[5]

The survey years we use are the most recent data by sector for which the young establishment identifiers are recorded. We use the provided sample weights to construct population-level estimates of total establishments and employment by district and three-digit NIC industry. Employment is formally defined as "persons engaged" and includes working owners, family and casual labor, and salaried employees.

[5] The sampling frame for the organized sector depends on the business register, and a concern might exist that firms indirectly sample out if they select a size so as to avoid registration. As we lack panel data, we are unable to take corrections like Disney et al. (2003). Absent a correlation with one of our explanatory variables, this measurement will primarily be for our outcome variables and thus not bias our estimates. With respect to the explanatory variables, we study the same covariates with the unorganized sector as well. We do not observe evidence of this type of gaming behavior when comparing results for the two sectors.

Districts are administrative subdivisions of Indian states or territories. Currently there are approximately 630 districts spread across 35 states/union territories. We exclude districts with a population less than one million (based on 2001 census) or with fewer than 50 establishments sampled. We exclude these small districts because limited sampling makes the data of limited value for our study (given that we need to evaluate district-industry conditions that separate young and incumbent establishments). We also exclude states that experienced ongoing conflict and political turmoil during the period of study. After these adjustments, the resulting sample retains districts in 20 major states that include more than 94% of Indian employment in both manufacturing and services.

Table 1 provides descriptive statistics. Figures 2 and 3 show spatial entry patterns, and the appendix offers additional tabulations by state. Entry rates, as a weighted average across all states, are 15% and 12% for organized and unorganized manufacturing, respectively. The entry rate is 20% for organized services, and 17% for unorganized services. The spatial entry rates for organized and unorganized sectors have -0.2 and 0.3 correlations across states for manufacturing and services, respectively. van Stel et al. (2007) emphasize the need to measure entry determinants separately across different types of entrepreneurs.

III. Determinants of Entrepreneurship

We now describe the spatial and industrial factors that we use to predict entrepreneurship. We first consider general traits of the district that affect all entrepreneurs, regardless of industry. These traits include both baseline features that are long-standing and slow to adjust, like the population distribution, and factors that are more directly influenced by policy makers, such as education and infrastructure, recognizing that deep change in education and infrastructure also takes a long time to accomplish. Second, recent research stresses the central importance of heterogeneity across industries as well as regions for explaining start-up rates (e.g., Fritsch and Falck 2007, Glaeser and Kerr 2009). Our second category thus develops industry-specific conditions that yield this heterogeneity within regional experiences.

District-Level Conditions

Our initial explanatory measures focus on basic traits of districts. It is essential to understand the effect of local area traits on entrepreneurship, especially given the disproportionate degree to which entrepreneurs found businesses in their home areas (e.g., Figueiredo et al. 2002, Michelacci and Silva 2007). We first control for population to provide a natural baseline of economic activity (e.g., consumer markets, general availability of workers). We next consider the district's age structure, measured as the ratio of working age population to non-working age population, given that the propensity to start new firms changes over the lifetimes of individuals,

and the age structure of a region often connects to local entry rates (e.g., Bönte et al. 2009, Delfmann et al. 2013). The age profile is often called the demographic dividend in the Indian context.

Third, we include a measure of population density. Unlike the clear positive predictions for the first two factors, the prediction for population density is ambiguous as it brings higher wages and land rents alongside greater market opportunities. Density has also been linked to stronger knowledge flows, and Audretsch and Fritsch (1994) use density as one source of convexity in local production that links to entry rates. Many studies link higher population density to reduced manufacturing entry rates, especially for larger plants that use established production techniques and seek to minimize costs. Duranton and Puga (2001) provide a formal theoretical model of this process. Ultimately, these multiple forces suggest an uncertain theoretical role for population density in explaining Indian entry rates.

Beyond these basic demographics, we consider five primary traits of districts: education of the local labor force, quality of local physical infrastructure, access or travel time to major Indian cities, stringency of labor laws, and household banking conditions. These traits are motivated by theoretical models of entrepreneurship and perceived importance to India's development, and we discuss other traits and their relationships to these variables below. Unless otherwise noted, these traits are taken from the 2001 Population Census.

Several studies link entrepreneurship to educated workforces in the U.S. (e.g., Doms et al. 2010, Glaeser et al. 2010), often with the underlying conceptual model that entrepreneurship requires a degree of creativity and handling of many tasks and ambiguous circumstances that education prepares one for (e.g., the model of Lazear 2005). Entrepreneurs may also benefit from specific development of basic business skills. On the other hand, Reynolds et al. (1994) do not find this relationship holds within every country, and Glaeser and Kerr (2009) find limited evidence for a link of education to U.S. manufacturing entrepreneurship. Thus, the literature is again ambiguous. Clarifying education's role for India is very important, as many local policy makers stress developing the human capital of their workforces, and India is no different (Amin and Mattoo 2008). We measure the general education level of a district by the percentage of adults with a graduate (post-secondary) degree. Our results are robust to alternative definitions such as the percentage of adults with higher secondary education.

Our second trait is the physical infrastructure level of the district. Basic services like electricity are essential for all businesses, but new entrants can be particularly dependent upon local infrastructure (e.g., established firms are better able to provision their own electricity if need be, which is quite common in India). Aghion et al. (2012) provide a recent theoretical model. Entrepreneurship is likely to benefit from greater infrastructure so long as the tax burden imposed to provide the infrastructure is not too high. Many observers cite upgrading India's

infrastructure as a critical step towards economic growth, and the Indian government has set aside substantial investment funds. The population census documents the number of villages in a district with telecommunications access, electricity access, paved roads, and safe drinking water. We calculate the percentage of villages that have infrastructure access within a district and sum across the four measures to create a continuous composite metric that ranges from zero (no infrastructure access) to four (full access).

India's economy is undergoing dramatic structural changes (Desmet et al. 2011). From a starting point in the 1980s when the government used licensing to promote industrial location in regions that were developing slowly, the economic geography of India has been in flux as firms and new entrants shift spatially (e.g., Fernandes and Sharma 2011). One feature that is important for a district in this transformation is its link to major cities. We thus include a measure from Lall et al. (2011) of the driving time to the nearest of India's ten largest cities as a measure of physical connectivity and across-district infrastructure.

We next model local labor regulations using state-level policy variation. Several studies link labor regulations in India to slower economic progress (e.g., Besley and Burgess 2004, Aghion et al. 2008), and Bozkaya and Kerr (2013) provide a theory model where tighter labor laws suppress entry. This effect may occur through reduced likelihood of wanting to start a new firm, or through reduced likelihood of opening new facilities from a desire to avoid regulations. There may also be reduced 'push' into entrepreneurship with more protected employment positions. We create a composite labor regulations index by state from the measures constructed by Ahsan and Pages (2007).

The final measure is the strength of household banking environment, reflecting the large literature on financial constraints and entrepreneurship, with Evans and Jovanovic (1989) being a seminal model. We measure the percentage of households that have banking services by district. This measure is likely to be particularly reflective of financing environments for unorganized sector activity.

Local Industrial Traits

Recent research emphasizes how local entrepreneurship varies substantially across industries, and our second set of metrics quantifies how suitable the local industrial environment is for a particular industry. Our first trait is the overall employment in a district-industry for incumbent firms. This is important given that entrepreneurs often leave incumbents to start their companies (e.g., Klepper 2010, Falck et al. 2008). From this baseline, we further develop metrics that unite the broad distribution of industry employments in districts with the extent to which industries interact through the traditional agglomeration rationales (e.g., Marshall 1920, Duranton and Puga 2004, Rosenthal and Strange 2004). We consider these forces within the manufacturing sector,

and we model these conditions through incumbent firms that pre-date the birth of the young businesses that we model in our outcome variables.[6]

The first agglomeration rationale is that proximity to customers and suppliers reduces transportation costs and thereby increases productivity. This reduction in shipping costs is the core agglomerative force of the New Economic Geography theory (e.g., Fujita et al. 1999). Where customers and suppliers are geographically separate, firms trade off distances. We measure the extent to which districts contain potential customers and suppliers for a new entrepreneur. We begin with an input-output table for India developed by the CSO. We define $Input_{i \leftarrow k}$ as the share of industry i's inputs that come from industry k, and $Output_{i \rightarrow k}$ as the share of industry i's outputs that go to industry k. These measures run from zero (no input or output purchasing relationship exists) to one (full dependency on the paired industry).

We summarize the quality of a district d in terms of its input flows for an industry i as $Input_{di} = - \sum_{k=1,...,I} abs(Input_{i \leftarrow k} - E_{dk}/E_d)$, where I indexes industries. This measure aggregates absolute deviations between the proportions of industrial inputs required by industry i and district d's actual industrial composition, with E representing employment. The measure is mostly orthogonal to district size, which we separately consider, and a negative value is taken so that the metric ranges between negative two (i.e., no inputs available) and zero (i.e., all inputs are available in the local market in precise proportions). This metric assumes that firms have limited ability to substitute across material inputs in their production processes.

To capture the relative strength of output relationships, we also define a consolidated metric $Output_{di} = \sum_{k=1,...,I} E_{dk}/E_d \cdot Output_{i \rightarrow k}$. This metric multiplies the national share of industry i's output sales that go to industry k with the fraction of industry k's employment in district d. By summing across industries, we take a weighted average of the strength of local industrial sales opportunities for industry i in the focal market d. This $Output_{di}$ measure takes on higher values with greater sales opportunities. This output measure allows greater substitution across customer industries than the design built into the input metric, and we have tested its robustness to several design variants.

Moving from material inputs, entrepreneurship is quite likely to be driven by the availability of a suitable labor force (e.g., the model of Combes and Duranton 2006). While education and demographics are informative about the suitability of the local labor force, these aggregate traits miss the very specialized nature of many occupations. Our working paper summarizes theories as to why specialized workers and firms agglomerate together and provides extended references. Unlike studies of advanced economies, India lacks the data to model direct occupational flows between industries. Greenstone et al. (2010) calculate from the Current

[6] This approach is used by Glaeser and Kerr (2009), Jofre-Monseny et al. (2011), Dauth (2011), and Mukim (2011). It follows upon the coagglomeration work of Ellison et al. (2010).

Population Survey the rate at which workers move between industries in the U.S. Using their measure of labor similarity for two industries, we define $Labor_{di}=\sum_{k=1,...,I} E_{dk}/E_d \cdot Mobility_{i \leftarrow k}$. This metric is a weighted average of the labor similarity of industries to the focal industry i, with the weights being each industry's share of employment in the local district. The metric is again by construction mostly orthogonal to district size.

These metrics condense large and diverse industrial structures for cities into manageable statistics of local industrial conditions. The advantages and limitations in their design are further discussed in our working paper. Perhaps the most important issue is that these district conditions do not capture interactions with neighboring districts, but factor and product markets can be wider than a local area. The average size of an Indian district is about the same as two U.S. counties at 5,500 square kilometers.

We finally turn to a special issue regarding local firm size distribution, building upon a literature that traces back to at least Johnson and Cathcart (1979). Fritsch and Falck (2007) and Parker (2009b) emphasize the strong degree to which an industrial base populated with small firms is associated with higher entrepreneurship rates. Fritsch and Falck (2007) note that the relationship could descend from a greater entrepreneurial culture[7], better training for entrepreneurs due to having worked in small businesses, or perhaps a reflection of the local industry's minimum efficient plant size. Parker (2009b) emphasizes a self-selection role by entrepreneurs. For the organized manufacturing sector, we test the inclusion of a measure of the local small firm share (<40 employees) in estimations. While there are many reasons to believe this pattern in advanced countries will carry over to India, there are also reasons to be doubtful. For example, Indian labor laws and size regulations have long suppressed average firm size in India compared to its peers, perhaps weakening this robust relationship evident elsewhere.

We also measure a specific variant of this effect related to customer/supplier industries. Chinitz (1961) observed that entrepreneurs often find it difficult to work with large, vertically-integrated suppliers. The entrepreneur's order sizes are too small, and often the entrepreneur's needs are non-standard. Empirical studies for the U.S. find the Chinitz effect very important in local start-up conditions. We quantify the Chinitz effect—as distinct from the general conditions captured in $Input_{di}$—through a metric that essentially calculates the average firm size in a district in industries that typically supply a given industry i: $Chinitz_{di}=\sum_{k=1,...,I} Firms_{dk}/E_d \cdot Input_{i \leftarrow k}$. Higher values of the $Chinitz_{di}$ metric indicate better supplier conditions for entrepreneurs.

[7] See Hofstede (2001), Boschma and Fritsch (2007), and Falck et al. (2011). Culture and social capital aspects are taken up by Fritsch and Wyrwich (2013), Kibler et al. (2013), and Westlund et al. (2013).

IV. Estimation Approach

We characterize factors related to entry through cross-sectional regressions at the district-industry level of India. This level of variation allows us to analyze both district-level determinants and the underlying heterogeneity for entrants across industries due to incumbent industrial structures. Following the above literature and conceptual notes, these specifications take the form

$$\ln(\text{Entry}_{di}) = \eta_i + \beta \cdot X_d + \gamma \cdot Z_{di} + \varepsilon_{di}.$$

The dependent variable is the log measure of entry employment by district-industry. Our sample includes the district-industry observations in which positive incumbent employment exists. The observation count thus differs across manufacturing and services and for organized and unorganized sectors. Many of our explanatory variables, such as incumbent district-industry employment, are also in log values so that the coefficients estimate proportionate responses. We transform non-log variables to have unit standard deviation for interpretation, weight estimations by an interaction of log industry size with log district population, and cluster standard errors by district to reflect the multiple mappings of district-level variables across local industries.[8]

We include a vector of industry fixed effects η_i in estimations. These fixed effects control for systematic differences across industries in their entrepreneurship rates, competition levels, average plant sizes, and similar. As Fritsch and Falck (2007) demonstrate, isolating spatial variations from these industry-level traits is very important. Also, our metrics of local industrial conditions utilize both fixed traits of industries (e.g., the input-output relationships, labor flows) and the distribution of industries within a district. The inclusion of industry fixed effects controls for these fixed industry-level traits except to the extent that they interact with the local industrial structure.

The vectors X_d and Z_{di} contain district and district-industry traits, respectively. Our estimation approach balances several objectives. First, given that there has been so little work on India, we seek to provide a sufficiently broad analysis to highlight where major correlations lie in the data. In doing so, we do not want to be too parsimonious in our specifications, but we also do not want to overload the analysis. Our set of metrics provides a good depiction of the Indian entrepreneurial landscape, motivated by theory, and we discuss in the robustness section and appendix the many additional factors we considered when forming this baseline.

[8] We recode a value of less than one entering employee on average as one entering employee. This maintains a consistent sample size, and the distinction between zero and one employee for a district-industry is not economically meaningful. These cells can be excluded without impacting our results.

We emphasize that this work measures partial correlations in the data, rather than causal parameters, reflective of our initial inquiry. In all cases, local traits are pre-determined for the entrepreneurship that we measure as the outcome variable. This provides some confidence against reverse causality, and we further test including lagged entry rates as a control variable. A second concern is omitted factors that are highly correlated with our regressors, making interpretation difficult. For example, in our baseline model, education may capture the quality of the local workforce that entrepreneurs employ, the strength of the local pool of potential entrepreneurs, and/or stronger local consumer demand. We provide some specific checks along these lines (e.g., controlling for consumption per capita), but there will be a natural limit against checking every feasible concern. We further discuss these issues below.

V. Empirical Results

Table 2A considers organized manufacturing. Column 1 includes just district populations, district-industry employments, and industry fixed effects. The existing district-industry employment strongly shapes the spatial location of entry: a 10% increase in incumbent employment raises entry employment by around 2%. In addition, a district's population increases entry rates with an elasticity of 0.5. Higher-order population terms are not found to be statistically significant or economically important. The adjusted R-Squared value for this estimation is quite modest at 0.13.

It is useful to compare these results to those evident in the U.S. for two reasons. First, the U.S.'s advanced economy—and policy environment that has relatively fewer distortions—provides a useful idea of what entrepreneurship and local conditions might look like at the frontier. This is not to say that India will necessarily look like the U.S. when it reaches current levels of U.S. development, just as entrepreneurship rates differ across advanced economies today. Nevertheless, in terms of broad regularities, it is very helpful to compare the India statistics against a country like the U.S. to provide perspective. A very well-known example in this regard is the Hsieh and Klenow (2009) comparison of misallocation of production across plants in India and the U.S. Second, and from an academic perspective, we have a growing body of evidence and intuition on how the U.S. economy functions with respect to entrepreneurship. The extent to which our study can identify where the Indian experience resembles or differs from the U.S. experience provides a reasonable starting point for ascertaining which lessons from the U.S. studies can be applied to the Indian context. Our conclusions section describes some of these lessons that do or do not apply, and hopefully our paper provides a touchstone for identifying whether lessons from future studies made of the U.S. or other advanced economies should be taken into account when thinking about the Indian context.

Glaeser and Kerr (2009) estimate a related specification for the U.S. that uses long-term employment for a city-industry as the key explanatory variable. If we adjust our estimation to more closely match their technique, we obtain an elasticity of 0.8 that is very similar to their 0.7 elasticity. While this elasticity is comparable, the R-Squared value for this estimation remains quite modest at 0.29, much lower than the R-Squared value of 0.80 for Glaeser and Kerr (2009). There are likely several factors behind this lower explanatory power for India, including data differences, estimations at the district versus city level, and similar. These natural differences between the Indian and U.S. data limit perfect comparison, but we believe the datasets are sufficiently similar to make some basic inference. Most important, it is clear that many industries within India's manufacturing sector are at a much earlier development stage than those in the U.S., where the manufacturing sector is instead shrinking. Thus, while existing patterns of industrial activity explain the similarity of spatial distribution of entrepreneurship in India and the U.S., India has much more variation in outcomes, which we characterize further below. Fernandes and Sharma (2011) also study these variations with respect to policy deregulations. Kathuria (2011) provides a broader exploratory framework.

Column 2 includes the district-level traits. Three factors stand out as discouraging entrepreneurship in organized manufacturing: high population density, strict labor regulations, and greater distance to one of India's ten biggest cities. The first pattern has been observed in many settings and reflects large manufacturers seeking cheaper environments. The second pattern connects with earlier studies of India that argue strict labor laws reduce economic growth. These policies are associated with reduced entry even after conditioning on district-industry size. The final factor highlights that while manufacturers avoid the high costs of urban areas, they also avoid the most remote areas of India in favor of settings that are relatively near to large population centers, likely to access customers directly or to connect to shipping routes. On the other hand, the education of a district's workforce is linked to higher entry rates. The elasticity that is estimated here is stronger than that found in comparable U.S. estimations.

The third column introduces district-industry traits. The roles of input and output markets are exceptionally strong with 0.4-0.5 elasticities. Both the labor market and Chinitz measures have positive coefficients. The decline in the main effect of incumbent employment suggests that these four new metrics capture the positive effects of local clusters on entry. The fourth column shows quite similar results if we further control for consumption per capita, per the discussion in the prior section. This control, along with the population metrics, suggests that demand side factors are not solely responsible for the positive roles that we see for metrics like education. Column 5 finds similar results when examining the log count of entering establishments, with the Chinitz metric being more prominent. We return to this difference when analyzing the entrant size distribution.

Across these columns of Table 2A, the R-Squared value increases from 0.13 to almost 0.3. While still modest, this growth in explanatory power due to modeling regional conditions is more substantial than that evident in Glaeser and Kerr (2009) for the U.S. This pattern highlights the greater relative importance of existing district conditions relative to incumbent positioning for explaining entrepreneurship in India, which we return to in the conclusions.

Table 2B considers unorganized manufacturing, and several differences exist to Table 2A. First, local population plays a much greater role, with approximately unit elasticity. Entrepreneurship in the unorganized sector is much more proportionate to local market sizes than in the organized sector. This theme is also evident in the independence of entry from local population density or travel time to a major city, the stronger relationship of entry to the age profile of the district, and the higher R-Squared values in Columns 1 and 2. Unorganized manufacturing clearly conforms much more closely to the overall contours of India's economic geography than organized manufacturing does.

The other two district traits that are associated with strong entry rates are the strength of local, within-district physical infrastructure and the strength of local household banking environments. This contrasts with organized manufacturing entry, where education stands out. An intuitive explanation, which will also be reflected in the services estimations, is that these patterns and their differences reflect the factors on which each sector depends most. Organized manufacturing establishments have broader resources that reduce dependency on local infrastructure and household finance. Likewise, it is reasonable to believe that the unorganized sector depends less on educated workers than the organized sector. While intuitive, these results should be viewed as partial correlations until they can be rigorously confirmed in future research.

We again find evidence for agglomeration economies within the unorganized manufacturing sector. The framework is similar to Table 2A except that we do not consider the Chinitz effect, since by definition the unorganized sector is comprised of small firms. Partly as a consequence of this, the inputs metric is relatively stronger in these estimations. The initial gap in explanatory power between the organized and unorganized sectors that was evident in the first two columns is diminished in the complete estimations in Columns 3-5.

Table 3 considers organized and unorganized services entry. The contrast to organized manufacturing is again quite intriguing. First, overall district population is as important as it was for unorganized manufacturing, with elasticity greater than one. Similarly, the R-Squared value grows to 0.20 and 0.47 with just the parsimonious set of explanatory factors in Columns 1 and 5, respectively. The R-Squared value using the Glaeser and Kerr (2009) approach for organized services is 0.30. Also similar to unorganized manufacturing, population density and travel time to major cities are not important in the multivariate setting, while the district's age profile does contribute to higher entry levels.

To recap, education and infrastructure matter the most among district traits. Education is generally more important, with particular relevance to organized sectors. Physical infrastructure has particular relevance to the unorganized sectors of the economy. The strength of the household banking sector is again also very important in the unorganized sectors of the economy. These channels provide three of the main ways that policy makers can influence the spatial distribution of entry.

The role of the existing incumbent employment by district-industry for services is weak in Table 3, likely suggesting that Marshallian economies are weaker in services. Unreported estimations further model Marshallian interactions in the services sector similar to manufacturing. These results are also weak, at most suggesting a small role for labor market interactions. However, we hesitate to strongly interpret this difference as the weak results may be due to applying concepts and metrics originally designed for manufacturers to the service sector.

Table 4 provides some extensions for organized manufacturing. Following the discussion in Section III, we first include in Column 1 the small firm incumbent share control. Including this control sharpens our earlier results further, including making the Chinitz effect more robust. We thus find evidence for the general small firm effects outline by Fritsch and Falck (2007) and Parker (2009b), as well as the Chinitz effect.

Columns 2-5 of Table 4 break out entrants by their sizes, and Table 5 provides a broader depiction of the entrant size distribution. Starting with Table 5, Panel A presents the full entrant distribution that includes the organized and unorganized sectors. The complete distribution across both sectors looks broadly similar to other environments. For example, 98% of entering establishments have fewer than 10 employees, and only 0.09% of entering establishments have more than 100 workers. In terms of employment shares, 76% of employment in entering establishments is contained in establishments with fewer than ten employees, versus 9.5% in those entering with more than 100. Panel B isolates the organized sector, and within this group the largest entrant size category contains 5.5% of establishments and 53% of employment. The district-level variation is also consistent around these traits.

Thus, the unorganized sector accounts for most entrants and employments, and includes plants that are by definition very small. The larger plants included in the organized sector are still skewed towards the smaller end of the size distribution (e.g., 79% have 10-19 employees) but the largest plants with more than 100 employees have 53% employment share. The definitions of entrants discussed in Section II highlight that our data include new firm formation, but also some elements of new establishments opening in a district. The former dominate the unorganized sector, given its small establishment sizes, while the latter become increasingly important in the larger size categories of the entrant size distribution for the organized sector. This makes a

separation very useful, as household banking conditions, for example, may matter less for the organized sector than the labor laws present in India.

Returning to Table 4, the heterogeneity across the entrant size distribution is fascinating and confirms many underlying theories and intuitions advanced earlier. Small entrants in the organized sector follow existing populations much more, similar to the unorganized sector (shown in Table 2b), while larger entrants in the organized sector are less tied to local demand and avoid places with high population density. The small business and Chinitz effects are much more important for small entrants in the organized sector, while labor markets and industrial output conditions are more critical for large entrants. Column 6 shows fairly similar results when using one-year entrants, with the main differences being greater emphasis on local banking conditions than local input markets. Column 7 likewise displays broadly similar results when instead controlling for lagged entry rates.

The appendix provides additional robustness checks on these results: excluding sample weights, including additional covariates like the female population share and local religious affiliations (e.g., Mack et al. 2013), and clustering standard errors by state. We have also tested controls for a district's caste population (Iyer et al. 2011), conflict, trade levels, and general development levels (leading/lagging designations at the state and district level). These additional controls do not substantively affect the results presented, and we maintain the more parsimonious specification to mirror other work from outside of India. Our main specifications are also robust to controlling for incumbent firm counts or value added rather than employment. The appendix also provides additional work regarding the local industrial traits. We obtain similar results when we include district fixed effects in the estimations, or when we use changes in industrial conditions from 1989 to 2005 to partially address omitted variable bias concerns.

VI. Conclusions

Entrepreneurship can be an important factor for economic growth, and India has historically had low entry rates for the formation of new businesses. This condition is starting to improve, and further growth in effective entrepreneurship is an important stepping stone in India's continued development. This paper explores the spatial determinants of local entrepreneurship for Indian manufacturing and services. This analysis provides an important baseline for understanding what is important in India's developing economy, both as a first step for policy advice and as a guide

to additional research efforts.[9] This foundation also serves a broader academic interest of comparing India's patterns with those of other economies like the U.S.

At the district level, our strongest evidence points to the roles that local education levels and physical infrastructure quality play in promoting entry. We also find evidence that strict labor regulations discourage entrepreneurship, and better household banking environments are associated with higher entry in the unorganized sector. Policy makers wishing to encourage entrepreneurship in their local areas have several policy levers that can be exploited: investment in both people and places is an easy call for policy makers, while reducing unnecessary regulations and restrictions is also warranted. This raises the importance of correct policy design for local areas, and it provides a nice testing ground for future work on agglomeration and urban economies. In particular, further research surrounding the time dimensions to entrepreneurship's role in the local economy (e.g., Fritsch and Mueller 2004) for India might be particularly attractive given the rapid pace of the country's transformation.

Research in regional science has also stressed the heterogeneity in entry across industries within a local area. We too find extensive evidence that the incumbent compositions of local industries influence new entry rates at the district-industry level within manufacturing. This influence is through traditional Marshallian agglomeration economies, the small firm effect that has been observed in many countries, and the Chinitz (1961) effect that emphasizes small suppliers. This evidence on localized agglomeration economies and entry is among the first in a developing economy of which we are aware.

Moving to comparative reflections, the similarities between the patterns observed for India and those in the U.S. are surprisingly large. For example, we were surprised at the strength of the small firm and Chinitz effects, given that many accounts of India describe how its firm size distribution has been artificially compressed. One could imagine the positive channels for entrepreneurship described for advanced economies being greatly diminished when the size distribution is being partially set by the government. Yet, these patterns are comparable. This general comparability is very important as it suggests a substantial degree of portability in the insights we have derived in studying advanced economies (e.g., Chatterji et al. 2013) to developing and emerging situations.[10]

The differences in the patterns between India and the U.S. are also instructive and provide important caveats and boundaries on this portability. First, the role and importance of education and physical infrastructure are higher in India than in comparable U.S. studies. By

[9] For example, Ghani et al. (2012) extend the distance to major city work by considering the development of the Golden Quadrangle highway system in India and its impact on districts (e.g., Datta 2011).

[10] Rosenthal and Strange (2013) and Ghani et al. (2013) identify similar features between the U.S. and India in the spatial sorting patterns of female entrepreneurs.

contrast, other dimensions like population density and regional age structures behave very similarly. Our conjecture is that the spatial variation in the latter dimensions within India more closely resembles the variation in advanced economies, and so the same underlying economic forces operate comparably. On the other hand, many parts of India struggle with illiteracy and lack of paved roads, which are not issues that regional comparisons from the U.S. can provide insights on. So, the important nuance to the broad comparability and portability noted earlier is that researchers and policy makers need to carefully contemplate whether the variations utilized in earlier studies are reflective of the variations that they are dealing with.

A second point of comparison with the U.S. is very striking. While coefficient elasticities are often similar in magnitude, a very striking difference between our work and that of Glaeser and Kerr (2009) is that we can generally only account for about a third of the spatial variation that the U.S.-focused study could. We posit that a large portion of this gap is due to India being at a much earlier stage of development, especially with the industrial landscape still adjusting to the deregulations of the 1980s and 1990s (e.g., Fernandes and Sharma 2011). District traits and local conditions take on a much greater importance, vis-à-vis incumbent employment distributions, with the economy in transition. At such an early point and with industrial structures not entrenched, local policies and traits can have profound and lasting impacts by shaping where industries plant their roots. These key differences between developing and advanced economies are worthy of further research.

References

Acs, Zoltan, and Attila Varga, "Entrepreneurship, Agglomeration and Technological Change", Small Business Economics 24:3 (2005), 323-334.

Aghion, Philippe, Robin Burgess, Stephen Redding, and Fabrizio Zilibotti, "The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India", American Economic Review 98 (2008), 1397-1412.

Aghion, Philippe, Ufuk Akcigit, Julia Cage, and William Kerr, "Taxation, Corruption, and Growth", Working Paper (2012).

Ahsan, Ahmed, and Carmen Pages, "Are All Labor Regulations Equal? Assessing The Effects Of Job Security, Labor Dispute And Contract Labor Laws In India", World Bank Working Paper 4259 (2007).

Amin, Mohammad, and Aaditya Mattoo, "Human Capital and the Changing Structure of the Indian Economy", World Bank Working Paper 4576 (2008).

Ardagna, Silvia, and Annamaria Lusardi, "Explaining International Differences in Entrepreneurship: The Role of Individual Characteristics and Regulatory Constraints", Working Paper (2008).

Audretsch, David, Maryann Feldman, Stephan Heblich, and Oliver Falck, "Local Entrepreneurship in Context", Regional Studies (2012), forthcoming.

Audretsch, David, and Michael Fritsch, "Growth Regimes over Time and Space", Regional Studies 36:2 (2002), 113-124.

Audretsch, David, and Michael Fritsch, "The Geography of Firm Births in Germany", Regional Studies 28:4 (1994), 359-365.

Baumol, William, "Entrepreneurship: Productive, Unproductive, and Destructive", Journal of Political Economy 98:5 (1990), 893-921.

Besley, Timothy, and Robin Burgess, "Can Labor Regulation Hinder Economic Performance? Evidence from India", Quarterly Journal of Economics 119 (2004), 91-134.

Bönte, Werner, Oliver Falck and Stephan Heblich, "The Impact of Regional Age Structure on Entrepreneurship", Economic Geography 85 (2009), 269-287.

Boschma, Ron, and Michael Fritsch, "Creative Class and Regional Growth − Empirical Evidence from Eight European Countries", Jena Working Paper 2007-066 (2007).

Bozkaya, Ant, and William Kerr, "Labor Regulations and European Venture Capital", Journal of Economics & Management Strategy (2013), forthcoming.

Calá, Carla, Joseph Maria Arauzo-Carod, and Miguel Manjón Antolín, "Regional Determinants of Firm Entry in a Developing Country", Working Paper (2013).

Chatterji, Aaron, Edward Glaeser, and William Kerr, "Clusters of Entrepreneurship and Innovation", NBER Working Paper 19013 (2013).

Chinitz, Benjamin, "Contrasts in Agglomeration: New York and Pittsburgh", American Economic Review 51:2 (1961), 279-289.

Combes, Pierre-Philippe, and Gilles Duranton, "Labour Pooling, Labour Poaching, and Spatial Clustering", Regional Science and Urban Economics 36 (2006), 1-28.

Datta, Saugato, "The Impact of Improved Highways on Indian Firms", Journal of Development Economics 99:1 (2011), 46-57.

Dauth, Wolfgang, "The Mysteries of the Trade: Interindustry Spillovers in Cities", Working Paper (2011).

Dehejia, Rajeev, and Arvind Panagariya, "Services Growth in India: A Look Inside the Black Box", Working Paper (2010).

Deichmann, Uwe, Somik Lall, Stephen Redding, and Anthony Venables, "Industrial Location in Developing Countries", World Bank Observer 23:2 (2008), 219-246.

Delfmann, Heike, Sierdjan Koster, Philip McCann, and Jouke van Dijk, "Population Change and New Firm Formation in Urban and Rural Regions", Working Paper (2013).

Delgado, Mercedes, Michael Porter, and Scott Stern, "Clusters and Entrepreneurship", Journal of Economic Geography 10:4 (2010), 495-518.

Desmet, Klaus, Ejaz Ghani, Stephen O'Connell, and Esteban Rossi-Hansberg, "The Spatial Development of India", Working Paper (2011).

Disney, Richard, Jonathan Haskel, and Ylva Heden, "Entry, Exit and Establishment Survival in UK Manufacturing", Journal of Industrial Economics 51:1 (2003), 91-112.

Doms, Mark, Ethan Lewis, and Alicia Robb, "Local Labor Force Education, New Business Characteristics, and Firm Performance", Journal of Urban Economics 67:1 (2010).

Drucker, Joshua, and Edward Feser, "Regional Industrial Dominance, Agglomeration Economies, and Manufacturing Plant Productivity", CES Working Paper 07-31 (2007).

Drucker, Joshua, and Edward Feser, "Regional Industrial Structure and Agglomeration Economies: An Analysis of Productivity in Three Manufacturing Industries", Regional Science and Urban Economics 42:1-2 (2012), 1-14.

Duranton, Gilles, and Diego Puga, "Micro-Foundations of Urban Agglomeration Economies", in Vernon Henderson and Jacques François Thisse (eds.) Handbook of Regional and Urban Economics, Volume 4 (Amsterdam: North-Holland, 2004), 2063–2117.

Duranton, Gilles, and Diego Puga, "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products", American Economic Review 91:5 (2001), 1454-1477.

Ellison, Glenn, Edward Glaeser, and William Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns", American Economic Review 100 (2010), 1195-1213.

Evans, David, and Boyan Jovanovic, "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints", Journal of Political Economy 97:4 (1989), 808-827.

Falck, Oliver, Michael Fritsch, and Stephan Heblich, "The Apple Does Not Fall From the Tree: Location of Start-Ups Relative to Incumbents", Jena Research Paper 2008-082 (2008).

Falck, Oliver, Michael Fritsch, and Stephan Heblich, "The Phantom of the Opera: Cultural Amenities, Human Capital, and Regional Economic Growth", Labour Economics 18:6 (2011), 755-766.

Fernandes, Ana, and Ariel Pakes, "Factor Utilization in Indian Manufacturing: A Look at the World Bank Investment Climate Survey Data", Working Paper (2010).

Fernandes, Ana, and Gunjan Sharma, "Together We Stand? Agglomeration in Indian Manufacturing", Working Paper (2011).

Figueiredo, Octávio, Paulo Guimaraes, and Douglas Woodward, "Home-Field Advantage: Location Decisions of Portuguese Entrepreneurs", Journal of Urban Economics 52:2 (2002), 341-361.

Fritsch, Michael, "How Does New Business Formation Affect Regional Development?", Small Business Economics 30 (2008), 1-14.

Fritsch, Michael, and Oliver Falck, "New Industry Formation by Industry over Space and Time: A Multidimensional Analysis", Regional Studies 41:2 (2007), 157-172.

Fritsch, Michael, and Pamela Mueller, "Effects of New Business Formation on Regional Development over Time", Regional Studies 38:8 (2004), 961-975.

Fritsch, Michael, and Michael Wyrwich, "The Long Persistence of Regional Entrepreneurship Culture: Germany 1925–2005", Jena Working Paper (2013).

Fujita, Masahisa, Paul Krugman, and Anthony Venables, The Spatial Economy: Cities, Regions and International Trade (Cambridge: MIT Press, 1999).

Ghani, Ejaz, Arti Grover Goswami, and William Kerr, "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing", Working Paper (2012).

Ghani, Ejaz, William Kerr, and Stephen O'Connell, "Promoting Entrepreneurship, Growth, and Job Creation" in Ejaz Ghani (ed.) Reshaping Tomorrow (Oxford: Oxford University Press, 2011).

Ghani, Ejaz, William Kerr, and Stephen O'Connell, "Local Industrial Structures and Female Entrepreneurship in India", Journal of Economic Geography (2013), forthcoming.

Glaeser, Edward, Sari Pekkala Kerr, and William Kerr, "Entrepreneurship and Urban Growth: An Empirical Assessment with Historical Mines", Working Paper (2012).

Glaeser, Edward, and William Kerr, "Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain?", Journal of Economics and Management Strategy 18:3 (2009), 623-663.

Glaeser, Edward, William Kerr, and Giacomo Ponzetto, "Clusters of Entrepreneurship", Journal of Urban Economics 67:1 (2010), 150-168.

Greenstone, Michael, Richard Hornbeck and Enrico Moretti, "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings", Journal of Political Economy 118 (2010), 536-598.

Hasan, Rana, and Karl Jandoc, "The Distribution of Firm Size in India: What Can Survey Data Tell Us?", ADB Working Paper 213 (2010).

Hofstede, Geert, Culture and Organizations (London: HarperCollins, 2001).

Hsieh, Chang-Tai, and Peter Klenow, "Misallocation and Manufacturing TFP in China and India", Quarterly Journal of Economics 124 (2009), 1403-1448.

Iyer, Lakshmi, Tarun Khanna, and Ashutosh Varshney, "Caste and Entrepreneurship in India", Working Paper (2011).

Jofre-Monseny, Jordi, Raquel Marín-López, and Elisabet Viladecans-Marsal, "The Mechanisms of Agglomeration: Evidence from the Effect of Inter-Industry Relations on the Location of New Firms", Journal of Urban Economics, forthcoming (2011).

Johnson, Peter, and D.G. Cathcart, "The Founders of New Manufacturing Firms: A Note on the Size of their 'Incubator's Plants'", Journal of Industrial Economics 28:2 (1979), 219-224.

Kathuria, Vinish, "What Causes Agglomeration? Policy or Infrastructure–A Study of Indian Manufacturing Industry", Working Paper (2011).

Kathuria, Vinish, Seethamma Natarajan, Rajesh Raj, and Kunal Sen, "Organized versus Unorganized Manufacturing Performance in India in the Post-Reform Period", MPRA Working Paper 20317 (2010).

Kibler, Ewald, Teemu Kautonen, and Matthias Fink, "Regional Social Legitimacy of Entrepreneurship: Implications for Entrepreneurial Intention and Start-up Behaviour", Working Paper (2013).

Khanna, Tarun, Billions of Entrepreneurs: How China and India Are Reshaping Their Futures—and Yours (Boston: Harvard University Press, 2008).

Klepper, Steven, "The Origin and Growth of Industry Clusters: The Making of Silicon Valley and Detroit", Journal of Urban Economics 67 (2010), 15-32.

Lall, Somik, Hyoung Wang, and Uwe Deichmann, "Infrastructure and City Competitiveness in India", Working Paper (2011).

Lazear, Edward, "Entrepreneurship", Journal of Labor Economics 23:4 (2005), 649-680.

Mack, Elizabeth, Alessandra Faggian, and Kevin Stolarick, "Does Religion Stifle Entrepreneurial Activity?", Working Paper (2013).

Marshall, Alfred, Principles of Economics (London: MacMillan, 1920).

Michelacci, Claudio, and Olmo Silva, "Why So Many Local Entrepreneurs?", Review of Economics and Statistics 89:4 (2007), 615-633.

Mueller, Pamela, Andre van Stel, and David Storey, "The Effects of New Firm Formation on Regional Development Over Time: The Case of Great Britain", Small Business Economics 30 (2008), 59-71.

Mukim, Megha, "Industry and the Urge to Cluster: A Study of the Informal Sector in India", Working Paper (2011).

Nataraj, Shanthi, "The Impact of Trade Liberalization on Productivity and Firm Size: Evidence from India's Formal and Informal Manufacturing Sectors", Working Paper (2009).

Parker, Simon, The Economics of Entrepreneurship (Cambridge: Cambridge University Press, 2009a).

Parker, Simon, "Why Do Small Firms Produce the Entrepreneurs?", Journal of Socio-Economics 38:3 (2009b), 484-494.

Reynolds, Paul, David Storey, and Paul Westhead, "Regional Characteristics Affecting Entrepreneurship: A Cross-National Comparison" in Frontiers of Entrepreneurship Research: 1994 (Babson Park, MA: Babson College, 1994), 550-564.

Reynolds, Paul, David Storey, and Paul Westhead, "Cross-National Comparisons of the Variation in New Firm Formation Rates", Regional Studies 41:S1 (2007), S123-S136.

Rosenthal, Stuart, and William Strange, "Evidence on the Nature and Sources of Agglomeration Economies", in Vernon Henderson and Jacques François Thisse (eds.) Handbook of Regional and Urban Economics, Volume 4 (Amsterdam: North-Holland, 2004), 2119-2171.

Rosenthal, Stuart, and William Strange, "Small Establishments/Big Effects: Agglomeration, Industrial Organization and Entrepreneurship", in Glaeser, Edward (ed.) Agglomeration Economics (Chicago: University of Chicago Press, 2010).

Rosenthal, Stuart, and William Strange, "Female Entrepreneurship, Agglomeration, and a New Spatial Mismatch", Review of Economics and Statistics 94 (2012), 764-788.

Sharma, Siddharth, "Factor Immobility and Regional Inequality: Evidence from a Credit Shock in India", Working Paper (2005).

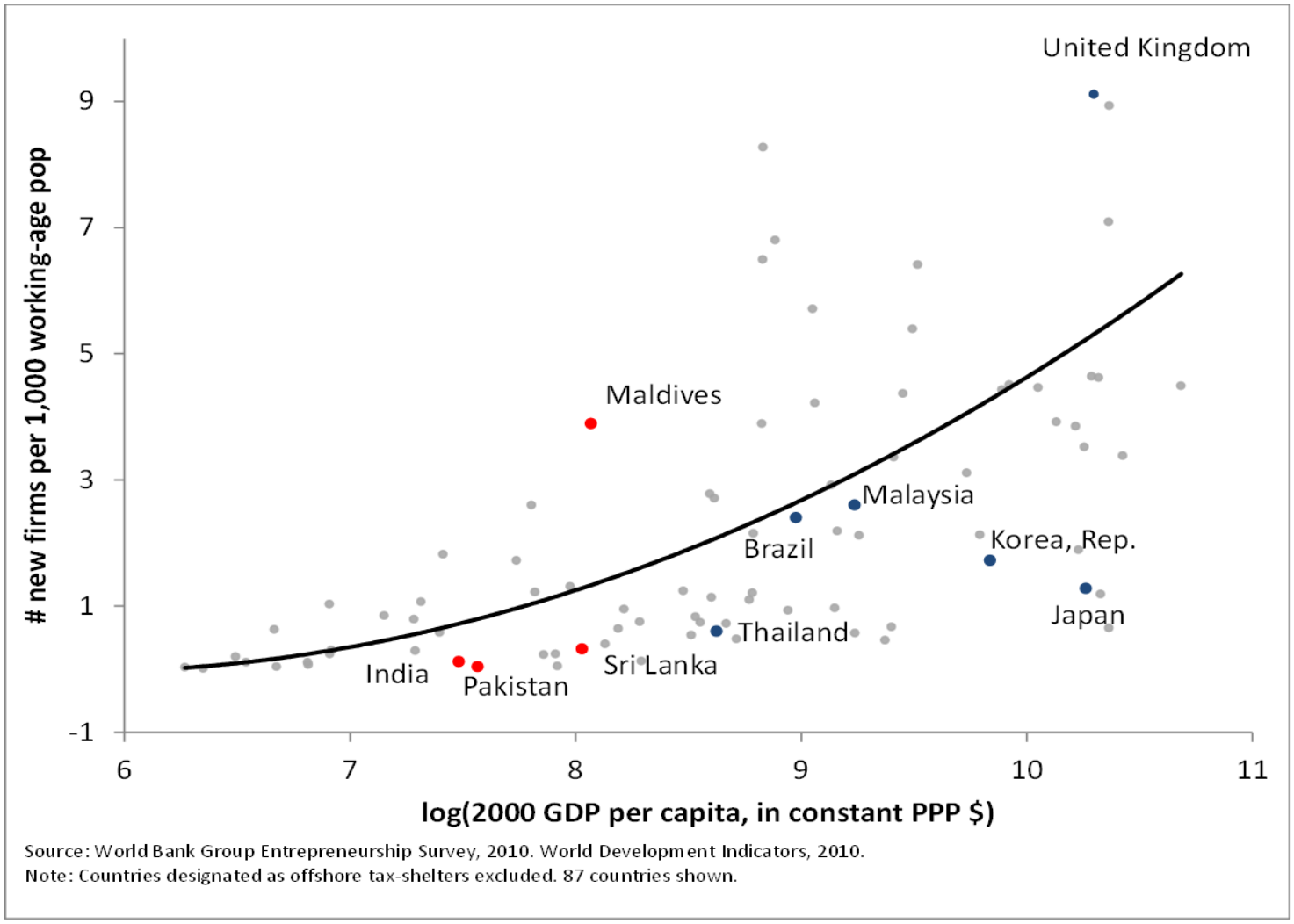
Storey, David, Understanding the Small Business Sector (Thomson Learning, 1994).

Storey, David, and Francis Greene, Small Business and Entrepreneurship (Prentice Hall, 2010).

van Stel, Andre, David Storey, and Roy Thurik, "The Effect of Business Regulations on Nascent and Young Business Entrepreneurship", Regional Studies 28 (2007), 171-186.

Westlund, Hans, Johan Larsson, and Amy Olsson, "Startups and Local Social Capital in the Municipalities of Sweden", Working Paper (2013).

Fig. 1: Business Registration Density, 2008



Source: World Bank Group Entrepreneurship Survey, 2010. World Development Indicators, 2010.
Note: Countries designated as offshore tax-shelters excluded. 87 countries shown.

Fig. 2: Indian Manufacturing Entry Rates, 2005-06

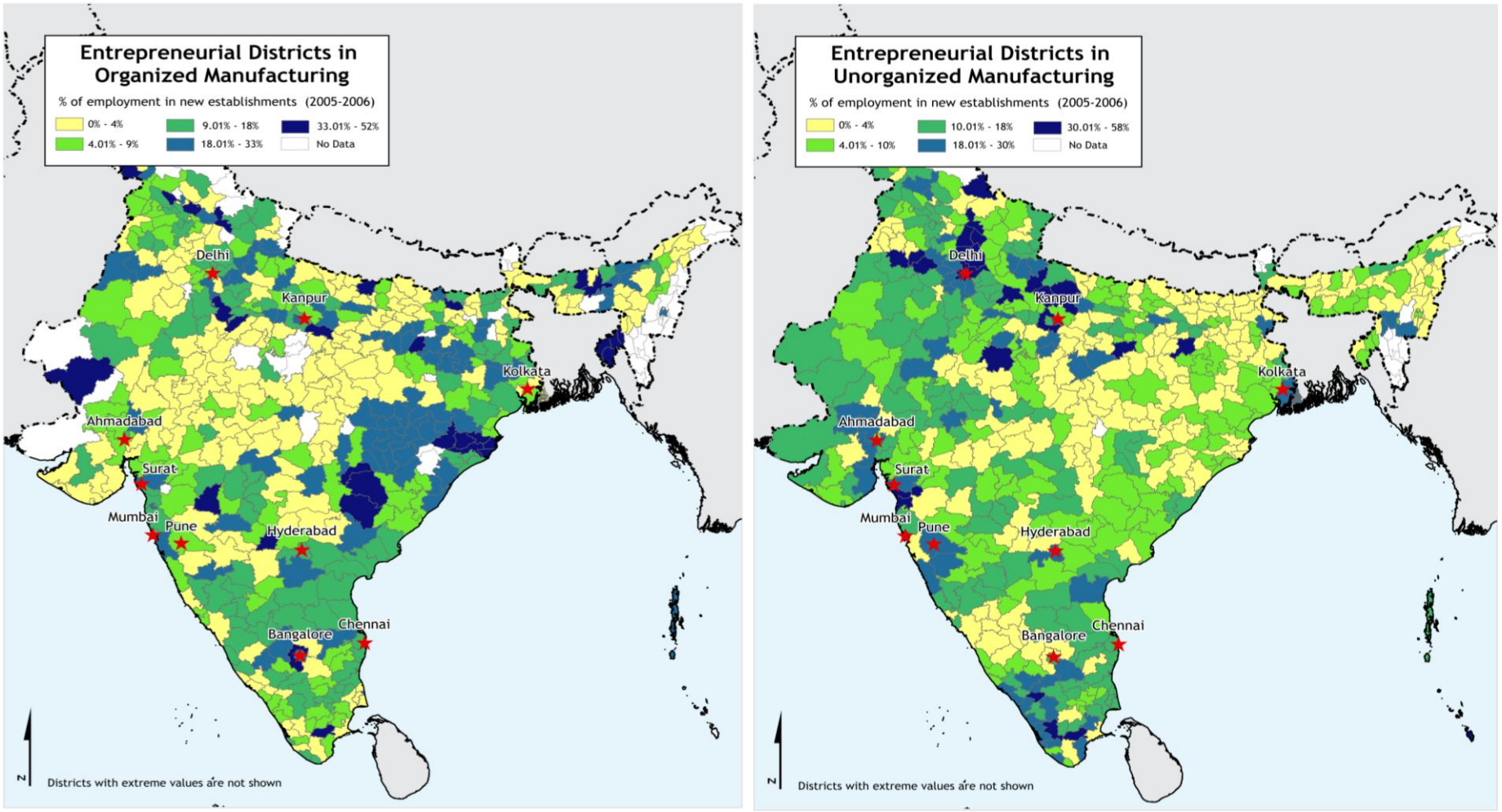
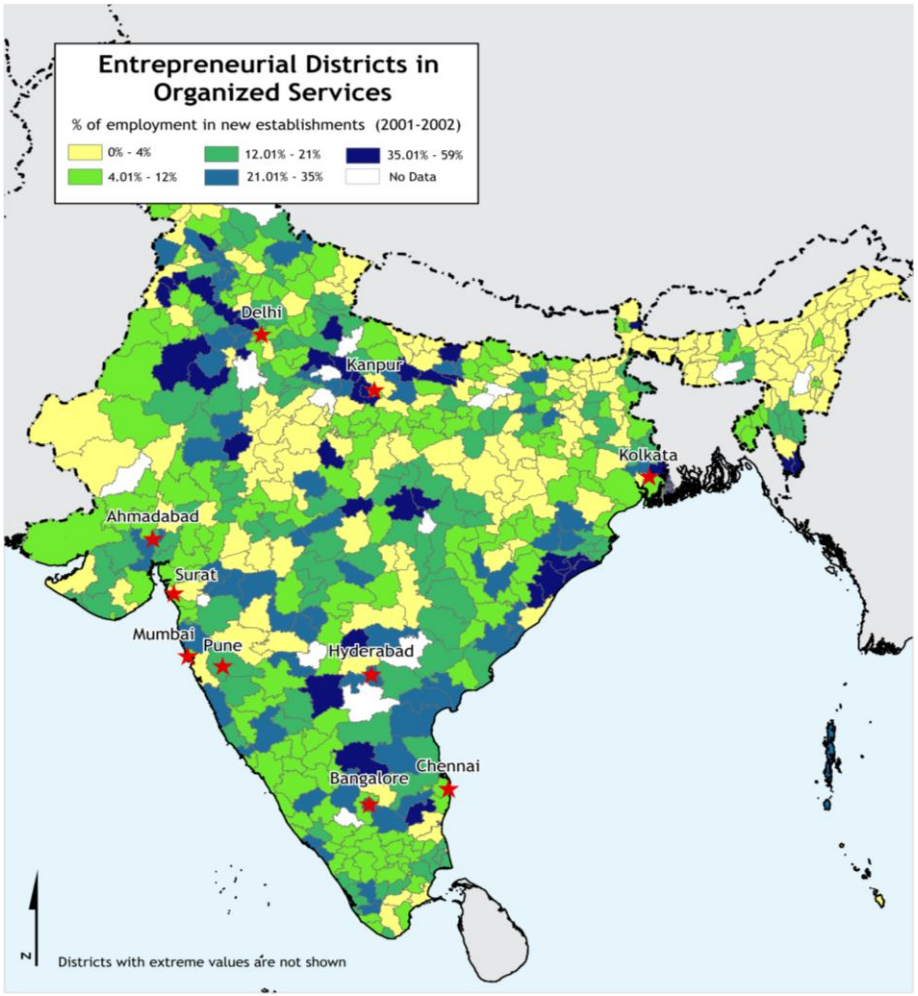


Fig. 3: Indian Services Entry Rates, 2001-02



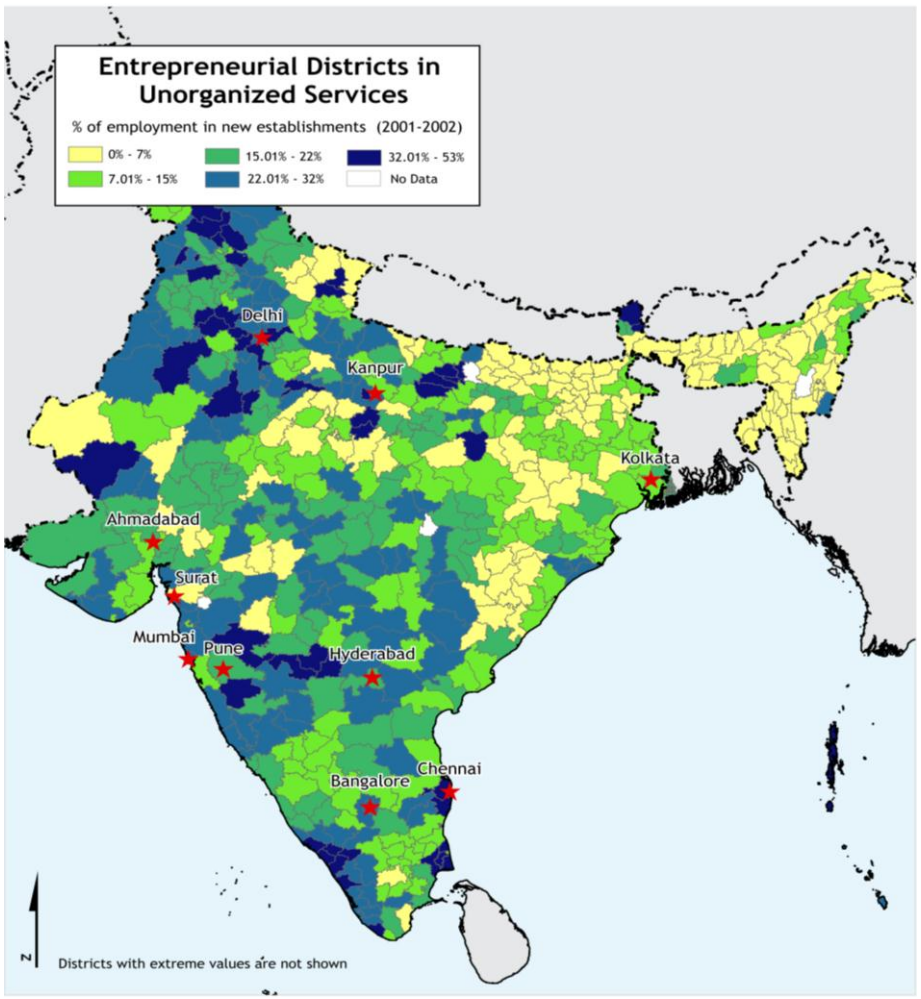


Table 1: Local industrial conditions for Indian entrepreneurship

	Mean	St. dev.	Median	Min	Max	Mean	St. dev.	Median	Min	Max
District traits:										
District population	2,972,828	1,731,997	3,207,232	1,021,573	13,900,000					
District population density (persons per sq. km.)	810	2,477	480	35	24,963					
Share of population with a graduate education	5.9%	2.7%	6.2%	1.7%	19.3%					
Demographic dividend for a district (age profile)	1.32	0.26	1.41	0.92	2.12					
Index of infrastructure quality for district	2.93	0.76	3.34	0.00	4.00					
Strength of household banking environment	0.35	0.13	0.38	0.09	0.73					
Stringency of labor laws: adjustments (state level)	0.69	0.84	0.00	0.00	3.00					
Stringency of labor laws: disputes (state level)	-0.41	1.24	0.00	-3.00	3.00					
Proximity to India's ten largest cities (min driving)	446	240	396	0	1,020					
Consumption per capita (2005USD at PPP)	680	186	625	352	1,397					
	Organized sector					Unorganized sector				
Industrial traits - Manufacturing:										
Total employment in district-industry	1,383	5,020	337	2	215,611	4,517	15,389	831	1	422,193
Start-up employment in district-industry	151	788	0	0	28,576	553	2,938	0	0	96,647
Labor market strength	0.09	0.13	0.05	0.00	0.97	0.09	0.11	0.04	0.00	0.97
Input / supplier strength	-1.64	0.25	-1.69	-2.00	-0.05	-1.71	0.24	-1.76	-2.00	-0.05
Chinitz index of small suppliers	0.48	1.33	0.25	0.00	45.52	n.a.	n.a.	n.a.	n.a.	n.a.
Output / customer strength (x10 for presentation)	0.01	0.02	0.03	0.00	7.64	0.01	0.02	0.00	0.00	0.97
Industrial traits - Services:										
Total employment in district-industry	1,761	5,892	400	3	173,293	2,885	8,145	376	1	195,863
Start-up employment in district-industry	268	1,429	0	0	47,048	502	1,581	46	0	50,243

Notes: Descriptive statistics based on Annual Survey of Industries and National Sample Survey, various rounds.

Table 2A: District entrepreneurship estimations - Organized Manufacturing

	Base estimation	District traits	Full estimation	Adding consumption	Using log entry count
	(1)	(2)	(3)	(4)	(5)
	DV is log entry employment by district-industry				
Log of incumbent employment in district-industry	0.229+++	0.186+++	-0.028	-0.030	0.032+
	(0.043)	(0.040)	(0.048)	(0.047)	(0.018)
Log of district population	0.531+++	0.483+++	0.475+++	0.482+++	0.216+++
	(0.179)	(0.155)	(0.156)	(0.161)	(0.056)
<u>District Traits:</u>					
Log of district population density		-0.569+++	-0.563+++	-0.562+++	-0.197+++
		(0.088)	(0.080)	(0.079)	(0.029)
Share of population with graduate education		0.211+	0.235++	0.230++	0.078+
		(0.110)	(0.107)	(0.111)	(0.042)
Demographic dividend for district (age profiles)		0.605	0.567	0.535	0.271
		(0.458)	(0.446)	(0.468)	(0.177)
Index of infrastructure quality for district		0.018	0.096	0.086	0.015
		(0.100)	(0.094)	(0.097)	(0.038)
Strength of household banking environment		0.143	0.095	0.085	0.027
		(0.104)	(0.100)	(0.106)	(0.036)
Stringency of labor laws in district's state		-0.210+++	-0.161++	-0.157++	-0.095+++
		(0.070)	(0.064)	(0.065)	(0.023)
Log travel time to closest large city		-0.275+++	-0.241+++	-0.237+++	-0.091+++
		(0.090)	(0.083)	(0.083)	(0.031)
Log per capita consumption				0.152	
				(0.505)	
<u>Local Industrial Conditions by Incumbent Firms:</u>					
Labor market strength for district-industry			0.161	0.164	0.026
			(0.102)	(0.102)	(0.041)
Inputs / supplier strength for district-industry			0.485+++	0.485+++	0.154+++
			(0.098)	(0.098)	(0.043)
Outputs / customer strength for district-industry			0.388+++	0.387+++	0.167+++
			(0.140)	(0.140)	(0.057)
Chinitz small suppliers metric for district-industry			0.279	0.279	0.337+++
			(0.213)	(0.212)	(0.129)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	4843	4843	4843	4843	4843
Adjusted R-squared	0.128	0.166	0.218	0.218	0.279

Notes: Estimations quantify the relationship between district-industry employment in new establishments and local conditions. District-level traits are taken from the 2001 Census. Industrial conditions are calculated from 2005-06 using incumbent establishments in the district-industry. Labor regulations are a composite of adjustment and disputes laws. Estimations weight observations by an interaction of district size and industry size, include industry fixed effects, and cluster standard errors by district. Non-logarithm variables are transformed to have unit standard deviation for interpretation.

Table 2B: District entrepreneurship estimations - Unorganized Manufacturing

	Base estimation	District traits	Full estimation	Adding consumption	Using log entry count
	(1)	(2)	(3)	(4)	(5)
	DV is log entry employment by district-industry				
Log of incumbent employment in district-industry	0.163+++	0.123+++	-0.075++	-0.078+++	-0.040
	(0.031)	(0.029)	(0.029)	(0.029)	(0.026)
Log of district population	1.051+++	0.878+++	1.010+++	1.025+++	0.866+++
	(0.161)	(0.157)	(0.160)	(0.153)	(0.138)
District Traits:					
Log of district population density		-0.019	-0.044	-0.042	-0.044
		(0.070)	(0.068)	(0.073)	(0.057)
Share of population with graduate education		-0.002	-0.026	-0.079	-0.046
		(0.080)	(0.084)	(0.087)	(0.074)
Demographic dividend for district (age profiles)		0.954+++	1.053+++	0.770++	0.798+++
		(0.326)	(0.330)	(0.326)	(0.285)
Index of infrastructure quality for district		0.386+++	0.365+++	0.259++	0.325+++
		(0.096)	(0.097)	(0.104)	(0.086)
Strength of household banking environment		0.222+++	0.211+++	0.152+	0.193+++
		(0.080)	(0.080)	(0.082)	(0.071)
Stringency of labor laws in district's state		-0.007	0.000	0.020	0.030
		(0.069)	(0.069)	(0.066)	(0.062)
Log travel time to closest large city		-0.004	0.009	0.029	0.017
		(0.069)	(0.074)	(0.074)	(0.065)
Log per capita consumption				1.191+++	
				(0.365)	
Local Industrial Conditions by Incumbent Firms:					
Labor market strength for district-industry			0.263+++	0.271+++	0.228+++
			(0.075)	(0.075)	(0.067)
Inputs / supplier strength for district-industry			0.553+++	0.542+++	0.504+++
			(0.107)	(0.108)	(0.096)
Outputs / customer strength for district-industry			0.291+++	0.292+++	0.246+++
			(0.050)	(0.051)	(0.044)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	6451	6451	6451	6451	6451
Adjusted R-squared	0.195	0.233	0.264	0.267	0.294

Notes: See Table 2A.

Table 3: District entrepreneurship estimations - Services

	Organized Services				Unorganized Services			
	Base estimation	District traits	Adding consumption	Using log entry count	Base estimation	District traits	Adding consumption	Using log entry count
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	DV is log entry employment by district-industry							
Log of incumbent employment in district-industry	-0.003	-0.104+++	-0.105+++	-0.054++	0.094+++	0.037+	0.037+	0.037+
	(0.038)	(0.033)	(0.033)	(0.023)	(0.024)	(0.021)	(0.021)	(0.021)
Log of district population	1.278+++	1.023+++	1.023+++	0.711+++	1.213+++	1.113+++	1.113+++	1.113+++
	(0.148)	(0.135)	(0.133)	(0.092)	(0.107)	(0.111)	(0.108)	(0.111)
Log of district population density		-0.014	-0.013	-0.028		-0.097+	-0.096+	-0.097+
		(0.086)	(0.087)	(0.056)		(0.057)	(0.058)	(0.057)
Share of population with graduate education		0.348+++	0.333+++	0.230+++		0.179+++	0.160++	0.179+++
		(0.085)	(0.088)	(0.059)		(0.068)	(0.070)	(0.068)
Demographic dividend for district (age profiles)		0.548+	0.469	0.329		0.574++	0.465++	0.574++
		(0.331)	(0.349)	(0.230)		(0.229)	(0.235)	(0.229)
Index of infrastructure quality for district		0.339+++	0.315+++	0.242+++		0.420+++	0.378+++	0.420+++
		(0.096)	(0.106)	(0.067)		(0.068)	(0.074)	(0.068)
Strength of household banking environment		0.174++	0.159+	0.108+		0.323+++	0.302+++	0.323+++
		(0.087)	(0.088)	(0.060)		(0.068)	(0.069)	(0.068)
Stringency of labor laws in district's state		-0.117+	-0.112+	-0.076+		-0.154+++	-0.146+++	-0.154+++
		(0.067)	(0.067)	(0.046)		(0.048)	(0.048)	(0.048)
Log travel time to closest large city		-0.011	-0.007	-0.021		0.048	0.056	0.048
		(0.054)	(0.054)	(0.037)		(0.051)	(0.050)	(0.051)
Log per capita consumption			0.295				0.454	
			(0.369)				(0.291)	
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3340	3340	3340	3340	6552	6552	6552	6552
Adjusted R-squared	0.201	0.252	0.253	0.252	0.471	0.536	0.536	0.536

Notes: See Table 2A.

Table 4: Extended district entrepreneurship estimations - Organized Manufacturing

	Incl. small firm share	Entering establishment employment of: 10-19	20-39	40-99	100+	One year entrants	Including lagged entry
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	DV is log entry employment by district-industry indicated in column header						
Log of incumbent employment in district-industry	0.534+++ (0.057)	0.229+++ (0.032)	0.277+++ (0.036)	0.260+++ (0.045)	0.274+++ (0.049)	0.407+++ (0.048)	-0.082+ (0.049)
Log of district population	0.358++ (0.143)	0.192++ (0.084)	0.266+++ (0.075)	0.178++ (0.085)	0.099 (0.099)	0.210 (0.133)	0.433+++ (0.152)
District Traits:							
Log of district population density	-0.453+++ (0.069)	-0.169+++ (0.038)	-0.160+++ (0.034)	-0.249+++ (0.043)	-0.281+++ (0.062)	-0.343+++ (0.053)	-0.521+++ (0.071)
Share of population with graduate education	0.229++ (0.099)	0.107+ (0.060)	0.089 (0.055)	0.066 (0.051)	0.118++ (0.054)	0.184++ (0.086)	0.235++ (0.106)
Demographic dividend for district (age profiles)	0.392 (0.410)	0.185 (0.255)	0.340 (0.240)	0.388 (0.244)	0.049 (0.309)	0.196 (0.335)	0.503 (0.449)
Index of infrastructure quality for district	0.011 (0.085)	-0.017 (0.061)	-0.042 (0.047)	-0.095 (0.058)	-0.033 (0.063)	-0.104 (0.070)	0.082 (0.086)
Strength of household banking environment	0.055 (0.085)	-0.002 (0.049)	0.017 (0.045)	0.058 (0.060)	0.090 (0.061)	0.187+++ (0.070)	0.061 (0.098)
Stringency of labor laws in district's state	-0.171+++ (0.060)	-0.094++ (0.037)	-0.145+++ (0.037)	-0.107+++ (0.038)	-0.036 (0.047)	-0.130++ (0.059)	-0.139++ (0.060)
Log travel time to closest large city	-0.183+++ (0.070)	-0.067 (0.041)	-0.064+ (0.035)	-0.121+++ (0.035)	-0.113++ (0.056)	-0.139++ (0.054)	-0.202++ (0.078)
Local Industrial Conditions by Incumbent Firms:							
Labor market strength for district-industry	0.034 (0.099)	-0.151++ (0.066)	-0.004 (0.068)	0.048 (0.074)	0.195++ (0.082)	-0.036 (0.087)	0.186+ (0.103)
Inputs / supplier strength for district-industry	0.204++ (0.086)	0.108+ (0.056)	0.064 (0.069)	0.049 (0.068)	0.059 (0.072)	0.050 (0.076)	0.429+++ (0.100)
Outputs / customer strength for district-industry	0.230++ (0.115)	0.111++ (0.053)	0.159++ (0.067)	0.247+++ (0.090)	0.275+++ (0.105)	0.235+++ (0.088)	0.364+++ (0.129)
Chinitz small suppliers metric for district-industry	0.429++ (0.209)	0.530+++ (0.184)	0.368++ (0.158)	0.150 (0.155)	-0.119 (0.139)	0.124 (0.156)	0.221 (0.214)
Share of small incumbent firms in the district-industry	0.651+++ (0.115)	0.447+++ (0.060)	0.409+++ (0.068)	0.254+++ (0.072)	0.055 (0.085)	0.169+++ (0.034)	
Lagged organized mfg entry rate for district-industry							0.205+++ (0.026)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4843	4843	4843	4843	4843	4843	4843
Adjusted R-squared	0.169	0.179	0.192	0.196	0.197	0.246	0.245

Notes: See Table 2a. Column 7 includes an unreported dummy variable for zero entry in the lagged period.

Table 5: Distribution of entrant employments and plant counts across size categories

	Employment in entering establishments			Counts of entering establishments		
	India as a whole	District-level mean	District-level stand. deviation	India as a whole	District-level mean	District-level stand. deviation
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Organized and unorganized sectors of manufacturing					
0-4 employees	59.11%	68.80%	26.85%	92.18%	92.79%	11.35%
5-9 employees	16.57%	10.46%	15.00%	5.83%	4.76%	9.67%
10-19 employees	8.85%	6.15%	10.15%	1.57%	1.65%	4.01%
20-39 employees	2.11%	2.34%	4.33%	0.18%	0.31%	0.78%
40-99 employees	3.83%	4.46%	8.82%	0.14%	0.33%	1.43%
100+ employees	9.52%	7.80%	14.01%	0.09%	0.16%	0.49%
	B. Organized sector only of manufacturing					
10-19 employees	36.42%	33.55%	32.63%	79.33%	57.81%	32.35%
20-39 employees	8.69%	14.39%	21.04%	9.18%	18.28%	22.00%
40-99 employees	15.74%	18.76%	23.30%	7.16%	14.69%	19.87%
100+ employees	39.15%	27.37%	29.93%	4.33%	9.23%	16.40%

Notes: Table documents the distribution of entrant employments and establishment counts across the establishment size distribution. The first and fourth columns provide statistics for India as a whole. The district mean and standard deviation columns summarize the unweighted variation at the district level.

Unpublished Appendix

App. Table 1 documents our data sources and years for each variable. For the infrastructure index, the underlying components are not reported in the Census data for six districts (major cities) that are not further subdivided into separate geographic units. In these cases we assign the infrastructure access components as 100%, and our results are robust to excluding these cities instead. The ten largest cities for the travel time calculations are Ahmedabad, Bangalore, Bhubaneshwar, Chennai, Delhi, Guwahati, Hyderabad, Kolkata, Mumbai, and Patna.

We note in the text that the input metric is mostly orthogonal to district size. The metric is not perfectly orthogonal to district size to the degree to which larger districts have more independent economic zones than smaller zones. Thus, even if the very localized input conditions within a small and large district are similar for a start-up, the measured quality of input conditions will be less in the larger district as the input metric will sum over more economic zones. We thank Juan Alcacer for pointing this out. The output measure also carries with it some residual dependencies on district size.

App. Tables 2A-3B provide descriptive statistics by state. App. Table 2A describes the organized manufacturing sector. There are just over 14,000 young establishments in India's organized manufacturing sector in 2005. This reflects an entry rate of approximately 15%, using a weighted average across states, which varies spatially to a large degree. Among the larger states in terms of manufacturing employment, entry rates are highest in Uttar Pradesh, Andhra Pradesh, and Karnataka at 18%-22%. Within Uttar Pradesh, the most entrepreneurial districts (on this relative scale) are Dehradun, Fatehpur, Faizabad, and Nagar Hardwar. The most entrepreneurial districts in Karnataka are Bangalore-Rural, Tumkur, Bangalore-Urban, and Dakshina Kannada. While possessing smaller manufacturing bases, entry rates are also high in Himachal Pradesh and Orissa. The spatial variation in entry rates across states is of similar magnitude to the regional variation evident in advanced economies (e.g., Reynolds et al. 2007).

The unorganized manufacturing sector in App. Table 2B has far more new establishments in any given year—almost 1.9 million in 2005 for our sample—although the entry rate at 12% is lower than the organized sector. There is negative correlation of -0.2 between spatial entry rates for organized and unorganized sectors across states. High rates of unorganized entry are found in Delhi, Haryana, Kerala, and Pondicherry, while Bihar, Karnataka, and Orissa have among the lowest rates. These contrasts are even starker when using self-employment measures: for the fifteen districts with self-employment accounting for greater than 50% of total district employment, none have an organized sector entry rate above 5%.

In the organized services sector document in App. Table 3A, there are about 120,000 young establishments in 2001, representing an entry rate of 20%. The highest rates are evident in Andhra Pradesh and Karnataka, with a number of other states closely following with entry rates of 20%-25%. Gujarat has the lowest entry rate. State-level entry rates for organized services have

a 0.4 correlation to those in organized manufacturing. The largest entry levels in absolute terms occur in the unorganized services sector documented in App. Table 3B, with over 2.2 million establishments at a rate of 17%. Entry rates are particularly high in Kerala and Haryana. Unorganized and organized activities are more closely linked in services than in manufacturing with a spatial correlation across states of 0.3 for services.

App. Tables 4 and 5 provide some key robustness checks for the organized and unorganized sectors, respectively. We repeat the full estimation for easy reference. We then show that the results are very similar without the sample weights. Our labor regulations metric varies across states, rather than districts, and so we show its precision (as well as the other variables) is robust to clustering standard errors by state. We then include the shares of the district population that are Hindu, Muslim, Christian, and Sikh as controls. Finally, we include the female population share as a control. Across these specifications, our reported results hold well, and we discuss in the main text some of the other controls we considered.

App. Tables 6 and 7 report conditional estimations that focus just on district-industry variation. App. Table 6 estimates a conditional specification of the form,

$$\ln(\text{Entry}_{di}) = \eta_i + \delta_d + \gamma \cdot \ln(\text{Incumbent Employment}_{di}) + \gamma_I \cdot \text{Input}_{di} + \gamma_O \cdot \text{Output}_{di} + \gamma_L \cdot \text{Labor}_{di} + \gamma_C \cdot \text{Chinitz}_{di} + \varepsilon_{di}.$$

We now include a vector of district fixed effects δ_d that control for differences across districts that are common for all industries, for example Delhi's larger size. Specifications thus employ variation within districts and industries: How much of the unexplained district-industry variation in entrepreneurship can we explain through incumbent local conditions that are especially suitable for particular industries?

The first three columns are for the organized manufacturing sector, while the last two columns are for the unorganized manufacturing sector. We report robust standard errors reflecting the district-industry variation. The coefficient patterns are stronger and more precisely estimated than those in the unconditional estimations reported in the main text. These results suggest that many of the agglomeration forces operate as strongly for entrants in India as they do in advanced economies, or perhaps even more strongly. We interpret the weaker performance in the paper's unconditional estimates, compared to these conditional estimates that fully control for district averages, as further evidence that India's economic geography is not set to the degree that an advanced economy is.

In addition to these conditional tests of our agglomeration metrics, unreported estimations confirm some of the district traits analyzed through interaction regressions. After including the district effects, we can no longer estimate the direct impact of labor laws on entry rates, but we can estimate an interaction of labor laws with how important labor is as an input factor for an industry. We estimate the latter importance through the industry's wage bill divided by industry value added. This interaction is negative and statistically significant, indicating that

entrepreneurship in labor-intensive sectors is disproportionately reduced by strict labor laws (similar conclusions are reached using employment-based measures). We likewise find that the Chinitz effect and local input conditions matter more in materially intensive industries.[1]

App. Table 7 examines the entrant size distribution for the organized sector by separating our overall entry measures into establishment sizes of young firms. The entry of a ten-person establishment is presumably a different phenomenon than the entry of a new firm with hundreds of employees. We care more about larger entrants in certain contexts, for example when worrying about the determinants of robust local labor demand. On the other hand, the entry of small establishments may be a purer reflection of entrepreneurship and hence more intrinsically interesting. More generally, empirical evidence exists that small and large establishments agglomerate differently (e.g., Holmes and Stevens 2002, Duranton and Overman 2008), and it is useful to extend this description to entering firms.[2]

App. Table 7 finds interesting distributional effects that also provide intuitive confirmation of the economic forces proposed. Most strikingly, the importance of the Chinitz effect is concentrated among small entrants, similar to the main text, while the importance of overall output markets and labor spillovers grow with entrant size. For India, it appears that input cost factors are more influential in the location choices of small start-ups, while output conditions and labor markets are more important for large entrants.

App. Table 8 contains our final set of empirical results. Our work thus far has focused on the cross-sectional patterns of incumbent industrial structures and entry. By including district and industry fixed effects, we focus on within-district and within-industry variation for analysis. This approach thus guards against omitted factors that vary by district or by industry. Similarly, our focus on incumbent firms to explain new entrants mirrors Jofre-Monseny et al. (2011), taking the former to be pre-determined. As an alternative, Glaeser and Kerr (2009) use predicted spatial distributions of industries due to natural cost advantages to provide a measure of exogeneity.

Nonetheless, a concern persists that there are unique aspects of district-industries that may confound this relationship. To take a United States example, the automobile industry has been concentrated in Detroit for over a century. Over this span of time, localized entrepreneurship and

[1] As another robustness check on our metric design, we find similar results when winsorizing our metrics at their 5% and 95% levels to weaken the influence of extreme values. The main adjustment is that the output metric grows in relative importance.

[2] A second rationale exists for examining the entry size distribution. Better local conditions may foster a larger entry size for entrepreneurs due to factors like less uncertainty about local growth potential and faster assembly of key resources. However, metrics of average entry size confound this intensive margin adjustment with changes in the extensive margin of greater entry rates. Better local conditions may simultaneously foster greater entry by many small firms, which leads to an overall decline in average entry size. We feel it is more prudent to look at the distribution measure.

incumbent industrial structures will have jointly influenced each other, and many other factors that we do not model may have arisen (e.g., special political connections and support by Detroit for the automobile industry). These latter factors that are particular to an industrial cluster would not be captured by city and industry fixed effects, and yet these instances of highly agglomerated activity are very important for identification in the above estimations. The long history of the Indian government's involvement in local industrial policy accentuates these econometric concerns for our estimates.

One approach to help address these concerns is to use time-varying conditions in localized agglomeration and entry by district-industry. By looking across two points in time, district-industry fixed effects can be included in the estimations. These fixed effects control for long-run levels of incumbent industrial structures and entry, focusing on changes within each district-industry. Such an approach does not fully overcome potential biases, as there could be time-varying factors within district-industries that continue to confound the analysis. The empirical bar, however, is set much higher.

A challenge to implementing this approach in many settings is that industrial structures can be very stable over time, providing little variation to exploit. India's organized manufacturing setting provides a unique opportunity in this regard. Prior to the large-scale deregulations, spatial location decisions for firms were set to a large degree by the government, with the goal to promote general equality across regions. In the two decades since these restrictions were lifted, India's manufacturing has seen large changes in spatial locations and agglomerations (e.g., Fernandes and Sharma 2011).

These changes provide much greater longitudinal variation than could typically be exploited. Micro-data for India's organized manufacturing sector extend back to 1989. We prepare our metrics for 1989 similar to those used in 2005. We restrict our sample to district-industry observations present in both periods. App. Table 8 estimates a panel specification of the form:

$$\ln(\text{Entry}_{dit}) = \eta_{it} + \delta_{dt} + \pi_{di} + \gamma \cdot \ln(\text{Incumbent Employment}_{dit}) + \gamma_I \cdot \text{Input}_{dit} + \gamma_O \cdot \text{Output}_{dit} + \gamma_L \cdot \text{Labor}_{dit} + \varepsilon_{dit}.$$

We now include a vector of district-industry fixed effects π_{di} that control for fixed differences across district-industries; we also extend our earlier fixed effects to be district-year and industry-year controls. These specifications thus employ panel variation: how much of the growth in district-industry entrepreneurship can we explain through changes in incumbent local conditions that are especially suitable for particular industries? By including district-year and industry-year fixed effects, we measure this effect after controlling for general district and industry development between 1989 and 2005.

App. Table 8 provides strong confirmation for our basic findings. In the first column, growth in general incumbent employment over the 16 years is linked to higher entrepreneurship. The elasticity is half the size estimated in the cross-section work. The elasticity decline may indicate

that an upward bias existed in the cross-sectional estimations (e.g., due to an omitted factor), but is also possible that the differencing exacerbated measurement error and its downward bias. In the second column, we also find support for the Marshallian metrics related to input and output markets. The coefficients are slightly larger than in the cross-sectional estimations and precisely estimated. Interestingly, labor conditions do not find support in the panel setting.[3] Changes in the Chinitz metric yielded implausibly large coefficients values due to outliers, and we do not report them. Of our metrics, the Chinitz effect is the most sensitive due to how it embodies both the establishment size distribution and input-output exchanges. Its sensitivity is thus not very surprising.

Overall, the panel estimations support our core evidence on the link between entrepreneurship and local industrial conditions in India. We think that India's industrial past, and the government-led spatial allocation of industrial activity that is rapidly becoming undone, provides a very interesting laboratory for testing many features of agglomeration and urban economics that are difficult to disentangle in advanced economies with more stable economic geographies.

[3] One potential factor that may contribute to the greater sensitivity of the labor metric compared to input-output measures is that the labor metric builds upon worker flows in the United States which may differ substantially from India. The input-output metrics, by contrast, directly build upon India's input-output tables.

App. Table 1: Data sources and years

Variable	Data source	Year
Organized manufacturing entry and incumbent structures data	GOI Ministry of Statistics and Programme Implementation Annual Survey of Industries 2005-2006	2005-06
	GOI Ministry of Statistics and Programme Implementation Annual Survey of Industries 1989-1990	1989-90
Unorganized manufacturing entry and incumbent structures data	GOI National Sample Survey Organisation, Socio-Economic Survey 62nd Round: July 2005 – June 2006, Schedule 2.2: Manufacturing Enterprises	2005-06
Services entry and incumbent structures data	GOI National Sample Survey Organisation, Socio-Economic Survey 57th Round: July 2001 - June 2002, Schedule 2.345: Unorganised Services	2001-02
District population	Census of India, District-level tabulations	2001
District population density (persons per sq. km.)	Census of India, District-level tabulations	2001
Share of population with a graduate education	Census of India, District-level tabulations	2001
Demographic divided for a district (age profile)	Census of India, District-level tabulations	2001
Index of infrastructure quality for district	Census of India, District-level tabulations	2001
Strength of household banking environment	Census of India, District-level tabulations	2001
Stringency of labor adjustment laws for district's state	Ahsan & Pages (2009)	2001
Stringency of labor disputes laws for district's state	Ahsan & Pages (2009)	2001
Travel time to closest of 10 largest cities (by population), in driving minutes	Authors' calculation	n/a
Consumption per capita (2005USD at purchasing power parity)	GOI National Sample Survey Organisation, Socio-Economic Survey 55th Round: July 1999 – June 2000, Household Schedule 10: Employment and Unemployment	1999

App. Table 2A: Summary statistics for Organized Manufacturing (2005-06)

	Number of districts in sample	Number of establishments	Number of new establishments (<3 years old)	Number of persons engaged	Number of persons engaged in new establ.	Share of establishments that are new	Share of employment in new establ.
Andhra Pradesh	23	9,835	2,032	813,062	65,373	21%	8%
Bihar	39	2,396	510	137,875	19,121	21%	14%
Chandigarh	1	242	12	8,747	327	5%	4%
Delhi	1	2,413	158	109,699	9,225	7%	8%
Goa	2	472	65	34,798	2,448	14%	7%
Gujarat	18	11,216	1,428	697,196	83,337	13%	12%
Haryana	16	3,772	279	354,229	35,713	7%	10%
Himachal Pradesh	11	691	261	52,120	17,391	38%	33%
Karnataka	17	5,401	981	459,946	80,582	18%	18%
Kerala	14	3,622	300	245,274	11,891	8%	5%
Madhya Pradesh	44	3,325	419	242,135	20,745	13%	9%
Maharashtra	30	14,058	1,749	1,035,176	101,400	12%	10%
Orissa	13	1,498	354	104,619	23,484	24%	22%
Pondicherry	4	480	108	37,516	5,007	22%	13%
Punjab	10	7,302	616	410,178	27,099	8%	7%
Rajasthan	26	4,865	731	257,648	28,570	15%	11%
Tamil Nadu	21	12,834	1,902	1,032,308	100,686	15%	10%
Uttar Pradesh	58	9,126	1,989	636,359	112,996	22%	18%
West Bengal	17	4,847	403	440,758	28,467	8%	6%

Notes: Descriptive statistics based on Annual Survey of Industries and National Sample Survey, various rounds. State-level traits are calculated over districts included in estimation samples.

App. Table 2B: Summary statistics for Unorganized Manufacturing (2005-06)

	Number of districts in sample	Number of establishments	Number of new establishments (<3 years old)	Number of persons engaged	Number of persons engaged in new establ.	Share of establishments that are new	Share of employment in new establ.
Andhra Pradesh	23	1,429,916	150,789	2,582,039	233,198	11%	9%
Bihar	39	1,304,917	60,122	2,082,220	101,453	5%	5%
Chandigarh	1	1,318	175	2,427	189	13%	8%
Delhi	1	97,424	27,483	448,017	152,965	28%	34%
Goa	2	10,132	775	25,575	1,845	8%	7%
Gujarat	18	637,662	106,278	1,384,543	215,246	17%	16%
Haryana	16	224,098	46,957	481,043	87,316	21%	18%
Himachal Pradesh	11	102,675	16,056	135,317	22,944	16%	17%
Karnataka	17	937,626	57,736	1,798,046	114,712	6%	6%
Kerala	14	569,116	103,104	1,092,465	220,759	18%	20%
Madhya Pradesh	44	1,017,799	85,875	1,852,854	156,109	8%	8%
Maharashtra	30	1,105,928	156,352	2,505,209	376,932	14%	15%
Orissa	13	830,361	40,010	1,435,721	64,572	5%	4%
Pondicherry	4	13,328	3,451	40,641	12,906	26%	32%
Punjab	10	284,890	32,573	547,420	59,377	11%	11%
Rajasthan	26	605,522	68,804	1,029,463	119,245	11%	12%
Tamil Nadu	21	1,454,360	192,973	3,052,908	429,085	13%	14%
Uttar Pradesh	58	2,250,792	393,130	4,418,741	738,044	17%	17%
West Bengal	17	2,689,853	353,045	4,742,212	533,644	13%	11%

Notes: See App. Table 2A.

App. Table 3A: Summary statistics for Organized Services (2001-02)

	Number of districts in sample	Number of establishments	Number of new establishments (<3 years old)	Number of persons engaged	Number of persons engaged in new establ.	Share of establishments that are new	Share of employment in new establ.
Andhra Pradesh	23	50,203	16,649	472,339	115,668	33%	24%
Bihar	39	24,335	2,426	186,560	15,992	10%	9%
Chandigarh	1	1,951	233	18,528	2,315	12%	12%
Delhi	1	22,471	3,278	232,993	23,766	15%	10%
Goa	2	1,250	262	12,622	2,150	21%	17%
Gujarat	17	42,785	3,275	536,731	24,907	8%	5%
Haryana	16	13,366	2,614	149,652	19,345	20%	13%
Himachal Pradesh	12	2,623	563	25,317	3,731	21%	15%
Karnataka	17	42,364	11,963	386,475	72,577	28%	19%
Kerala	14	34,709	7,037	361,322	55,934	20%	15%
Madhya Pradesh	44	38,538	5,548	374,968	43,684	14%	12%
Maharashtra	30	90,819	19,160	869,617	160,806	21%	18%
Orissa	13	18,178	4,200	148,874	27,274	23%	18%
Pondicherry	4	2,762	669	29,876	7,527	24%	25%
Punjab	10	14,438	3,463	185,046	29,020	24%	16%
Rajasthan	26	26,970	6,491	244,282	48,074	24%	20%
Tamil Nadu	21	48,190	7,841	548,315	59,680	16%	11%
Uttar Pradesh	58	103,267	20,564	897,253	171,210	20%	19%
West Bengal	17	35,660	4,445	382,695	38,626	12%	10%

Notes: See App. Table 2A.

App. Table 3B: Summary statistics for Unorganized Services (2001-02)

	Number of districts in sample	Number of establishments	Number of new establishments (<3 years old)	Number of persons engaged	Number of persons engaged in new establ.	Share of establishments that are new	Share of employment in new establ.
Andhra Pradesh	23	1,283,845	227,385	2,056,671	352,485	18%	17%
Bihar	39	1,433,440	69,281	2,010,992	98,044	5%	5%
Chandigarh	1	16,092	3,731	25,496	5,607	23%	22%
Delhi	1	224,651	57,135	372,375	93,097	25%	25%
Goa	2	12,705	3,222	22,494	5,223	25%	23%
Gujarat	17	556,165	85,688	779,596	122,316	15%	16%
Haryana	16	208,836	61,600	299,033	85,968	29%	29%
Himachal Pradesh	12	77,258	24,386	116,024	36,810	32%	32%
Karnataka	17	633,884	121,182	996,150	200,301	19%	20%
Kerala	14	560,170	179,377	833,808	272,035	32%	33%
Madhya Pradesh	44	567,032	87,800	872,027	143,214	15%	16%
Maharashtra	30	1,094,885	252,756	1,732,570	416,709	23%	24%
Orissa	13	593,974	66,535	997,465	100,347	11%	10%
Pondicherry	4	16,095	3,887	32,017	8,068	24%	25%
Punjab	10	343,779	81,993	474,572	117,711	24%	25%
Rajasthan	26	514,407	107,974	748,729	161,132	21%	22%
Tamil Nadu	21	962,148	176,200	1,575,196	301,720	18%	19%
Uttar Pradesh	58	2,472,139	446,279	3,448,423	663,916	18%	19%
West Bengal	17	1,524,382	166,715	1,939,358	216,087	11%	11%

Notes: See App. Table 2A.

App. Table 4: Robustness checks on district entrepreneurship estimations - Organized Manufacturing and Services

	Manufacturing					Services				
	Full estimation	No weights	Clustering by state	Incl. religion controls	Incl. female pop share	Full estimation	No weights	Clustering by state	Incl. religion controls	Incl. female pop share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	DV is log entry employment by district-industry									
Log of incumbent employment in district-industry	-0.028	-0.075++	-0.028	-0.027	-0.033	-0.104+++	-0.098+++	-0.104+++	-0.104+++	-0.104+++
	(0.048)	(0.038)	(0.037)	(0.047)	(0.048)	(0.033)	(0.028)	(0.031)	(0.033)	(0.033)
Log of district population	0.475+++	0.414+++	0.475++	0.442+++	0.461+++	1.023+++	0.914+++	1.023+++	1.030+++	1.011+++
	(0.156)	(0.137)	(0.187)	(0.166)	(0.160)	(0.135)	(0.121)	(0.149)	(0.139)	(0.136)
District Traits:										
Log of district population density	-0.563+++	-0.493+++	-0.563+++	-0.548+++	-0.561+++	-0.014	0.019	-0.014	-0.012	-0.015
	(0.080)	(0.087)	(0.090)	(0.079)	(0.087)	(0.086)	(0.084)	(0.098)	(0.086)	(0.083)
Share of population with graduate education	0.235++	0.208++	0.235+	0.214+	0.163	0.348+++	0.299+++	0.348+++	0.346+++	0.311+++
	(0.107)	(0.096)	(0.120)	(0.109)	(0.111)	(0.085)	(0.081)	(0.054)	(0.083)	(0.087)
Demographic dividend for district (age profiles)	0.567	0.441	0.567	0.930+	0.931++	0.548+	0.534+	0.548+	0.536	0.698+
	(0.446)	(0.393)	(0.481)	(0.550)	(0.440)	(0.331)	(0.288)	(0.285)	(0.405)	(0.365)
Index of infrastructure quality for district	0.096	0.111	0.096	0.082	0.066	0.339+++	0.297+++	0.339++	0.331+++	0.327+++
	(0.094)	(0.083)	(0.107)	(0.097)	(0.092)	(0.096)	(0.080)	(0.122)	(0.102)	(0.098)
Strength of household banking environment	0.095	0.070	0.095	0.137	0.075	0.174++	0.173++	0.174++	0.173++	0.176++
	(0.100)	(0.089)	(0.144)	(0.099)	(0.098)	(0.087)	(0.078)	(0.082)	(0.088)	(0.087)
Stringency of labor laws in district's state	-0.161++	-0.139++	-0.161++	-0.198+++	-0.184+++	-0.117+	-0.118++	-0.117+	-0.114	-0.131+
	(0.064)	(0.056)	(0.061)	(0.060)	(0.062)	(0.067)	(0.058)	(0.065)	(0.074)	(0.068)
Log travel time to closest large city	-0.241+++	-0.225++	-0.241++	-0.212++	-0.216++	-0.011	-0.003	-0.011	-0.014	-0.002
	(0.083)	(0.087)	(0.099)	(0.087)	(0.083)	(0.054)	(0.053)	(0.043)	(0.055)	(0.055)
Local Industrial Conditions by Incumbent Firms:										
Labor market strength for district-industry	0.161	0.149	0.161+	0.170+	0.171+					
	(0.102)	(0.094)	(0.085)	(0.100)	(0.102)					
Inputs / supplier strength for district-industry	0.485+++	0.563+++	0.485+++	0.484+++	0.487+++					
	(0.098)	(0.088)	(0.074)	(0.097)	(0.098)					
Outputs / customer strength for district-industry	0.388+++	0.434+++	0.388+++	0.383+++	0.379+++					
	(0.140)	(0.148)	(0.130)	(0.140)	(0.137)					
Chinitz small suppliers metric for district-industry	0.279	0.139	0.279	0.278	0.259					
	(0.213)	(0.176)	(0.214)	(0.212)	(0.212)					
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4843	4843	4843	4843	4843	3340	3340	3340	3340	3340
Adjusted R-squared	0.218	0.194	0.218	0.220	0.220	0.252	0.240	0.252	0.252	0.253

Notes: Robustness checks on Table 2A and Table 3.

App. Table 5: Robustness checks on district entrepreneurship estimations - Unorganized Manufacturing and Services

	Manufacturing					Services				
	Full estimation	No weights	Clustering by state	Incl. religion controls	Incl. female pop share	Full estimation	No weights	Clustering by state	Incl. religion controls	Incl. female pop share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	DV is log entry employment by district-industry									
Log of incumbent employment in district-industry	-0.075++	-0.053++	-0.053+	-0.081+++	-0.073++	0.037+	0.047+++	0.037+	0.039+	0.037+
	(0.029)	(0.024)	(0.027)	(0.029)	(0.029)	(0.021)	(0.018)	(0.018)	(0.021)	(0.020)
Log of district population	1.010+++	0.853+++	0.853+++	0.942+++	1.004+++	1.113+++	0.977+++	1.113+++	1.101+++	1.112+++
	(0.160)	(0.131)	(0.138)	(0.164)	(0.160)	(0.111)	(0.100)	(0.099)	(0.113)	(0.111)
District Traits:										
Log of district population density	-0.044	-0.055	-0.055	-0.065	-0.045	-0.097+	-0.067	-0.097	-0.093	-0.097+
	(0.068)	(0.065)	(0.076)	(0.073)	(0.064)	(0.057)	(0.060)	(0.069)	(0.057)	(0.057)
Share of population with graduate education	-0.026	-0.019	-0.019	0.035	-0.091	0.179+++	0.170+++	0.179++	0.171++	0.177++
	(0.084)	(0.073)	(0.088)	(0.086)	(0.087)	(0.068)	(0.065)	(0.079)	(0.068)	(0.072)
Demographic dividend for district (age profiles)	1.053+++	0.942+++	0.942+	1.006+++	1.359+++	0.574++	0.533++	0.574+	0.699+++	0.585++
	(0.330)	(0.286)	(0.468)	(0.365)	(0.348)	(0.229)	(0.214)	(0.329)	(0.257)	(0.281)
Index of infrastructure quality for district	0.365+++	0.313+++	0.313+	0.388+++	0.339+++	0.420+++	0.391+++	0.420+++	0.405+++	0.419+++
	(0.097)	(0.083)	(0.153)	(0.100)	(0.097)	(0.068)	(0.061)	(0.093)	(0.071)	(0.068)
Strength of household banking environment	0.211+++	0.194+++	0.194++	0.161+	0.216+++	0.323+++	0.299+++	0.323+++	0.331+++	0.323+++
	(0.080)	(0.072)	(0.082)	(0.084)	(0.080)	(0.068)	(0.065)	(0.082)	(0.071)	(0.068)
Stringency of labor laws in district's state	0.000	-0.029	-0.029	0.008	-0.019	-0.154+++	-0.146+++	-0.154+++	-0.168+++	-0.154+++
	(0.069)	(0.059)	(0.097)	(0.071)	(0.071)	(0.048)	(0.044)	(0.034)	(0.052)	(0.048)
Log travel time to closest large city	0.009	0.002	0.002	0.021	0.028	0.048	0.056	0.048	0.053	0.049
	(0.074)	(0.066)	(0.098)	(0.074)	(0.074)	(0.051)	(0.047)	(0.056)	(0.052)	(0.053)
Local Industrial Conditions by Incumbent Firms:										
Labor market strength for district-industry	0.263+++	0.213+++	0.213++	0.262+++	0.266+++					
	(0.075)	(0.066)	(0.090)	(0.075)	(0.075)					
Inputs / supplier strength for district-industry	0.553+++	0.652+++	0.652+++	0.579+++	0.543+++					
	(0.107)	(0.097)	(0.114)	(0.107)	(0.107)					
Outputs / customer strength for district-industry	0.291+++	0.264+++	0.264+++	0.287+++	0.284+++					
	(0.050)	(0.055)	(0.069)	(0.051)	(0.050)					
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6451	6458	6458	6458	6458	6552	6557	6552	6552	6552
Adjusted R-squared	0.264	0.252	0.252	0.269	0.265	0.536	0.523	0.536	0.536	0.536

Notes: Robustness checks on Table 2B and Table 3.

App. Table 6: Conditional entrepreneurship estimations - Manufacturing

	Organized Manufacturing			Unorg. Manufacturing	
	Base estimation	Including Marshallian	Including Chinitz	Base estimation	Including Marshallian
	(1)	(2)	(3)	(4)	(5)
	DV is log entry employment by district-industry				
Log of incumbent employment in district-industry	0.155+++	-0.064++	-0.063++	0.109+++	-0.100+++
	(0.025)	(0.031)	(0.030)	(0.026)	(0.026)
Labor market strength for district-industry		0.515+++	0.496+++		0.370+++
		(0.089)	(0.089)		(0.066)
Inputs / supplier strength for district-industry		0.344+++	0.263+++		0.448+++
		(0.080)	(0.088)		(0.108)
Outputs / customer strength for district-industry		0.291+++	0.304+++		0.264+++
		(0.113)	(0.111)		(0.045)
Chinitz small suppliers metric for district-industry			0.597+++		
			(0.203)		
District and industry fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	4843	4843	4843	6451	6451
Adjusted R-squared	0.259	0.306	0.309	0.308	0.340

Notes: See Table 2A. Estimations consider entry patterns conditional on industry and district fixed effects.

App. Table 7: Entry size distribution estimations - Manufacturing

	Entering establishment employment of:			
	10-19	20-39	40-99	100+
	(1)	(2)	(3)	(4)
	DV is log entry employment by district-industry			
Log of incumbent employment in district-industry	0.038++	0.062+++	0.048++	0.024
	(0.016)	(0.019)	(0.022)	(0.027)
Labor market strength for district-industry	-0.002	0.134++	0.218+++	0.447+++
	(0.055)	(0.066)	(0.071)	(0.087)
Inputs / supplier strength for district-industry	0.106+	0.127++	0.115+	0.046
	(0.055)	(0.063)	(0.069)	(0.080)
Outputs / customer strength for district-industry	0.142+++	0.160++	0.226+++	0.277+++
	(0.051)	(0.064)	(0.079)	(0.091)
Chinitz small suppliers metric for district-industry	0.598+++	0.384++	0.108	-0.028
	(0.167)	(0.153)	(0.138)	(0.138)
District and industry fixed effects	Yes	Yes	Yes	Yes
Observations	4843	4843	4843	4843
Adjusted R-squared	0.190	0.188	0.194	0.219

Notes: See Table 2A and App. Table 6. Estimations disaggregate entry into an entry size distribution based upon initial employment in the establishment.

App. Table 8: Entrepreneurship estimations using changes from 1989 to 2005

	Base estimation	Including Marshallian
	(1)	(2)
	DV is log entry empl.	
Log of incumbent employment in district-industry	0.068++	-0.083++
	(0.033)	(0.036)
Labor market strength for district-industry		-0.085
		(0.127)
Inputs / supplier strength for district-industry		0.428+++
		(0.124)
Outputs / customer strength for district-industry		0.495+++
		(0.087)
District-industry, district-year, and industry-year fixed effects	Yes	Yes
Observations	5848	5848
Adjusted R-squared	0.269	0.292

Notes: See Table 2A and App. Table 6. Estimations consider changes in incumbent industrial conditions and entrepreneurship from 1989 to 2005 for the organized manufacturing sector.